Federation Proceedings

VOLUME 7 *September 1948* NUMBER 3

AMERICAN PHYSIOLOGICAL SOCIETY

Symposium on Physiology of Neuro-Muscular Junctions[1]

JOHN H. WELSH, *Chairman*

PHYSIOLOGY OF NEURO-MUSCULAR JUNCTIONS

JOHN H. WELSH

From the Biological Laboratories, Harvard University, Cambridge, Massachusetts

Since the contributions to this symposium are concerned mainly with the problem of junctional transmission between nerve and muscle in the vertebrates, it may be useful, by way of introduction, to sketch briefly the evolution of mechanisms of communication between cells. We often gain a better understanding of the biology of the vertebrates if we have some notion of the ways in which more primitive animals have achieved essentially the same ends.

Most of the protozoa have no problem of intercellular communication since they consist largely of single cells. They do, however, have the problem of intra-cellular coördination of their many activities and have obviously achieved a satisfactory solution to this problem. For example, the fibrillar system for the control of ciliary beat in ciliate protozoa is an effective means of coördinating the movements of these numerous organelles.

With the advent of multicellular animals the need arose for coördination of the activities of their many cells and organs. No longer could messages be conducted through continuous protoplasm and over relatively short distances, but cell membranes and long distances imposed new problems. One solution to the problems of distance and membrane barriers was to have certain cells develop elongated processes. This is seen for the first time in the coelenterates where sensory cells are found with long processes which end in close association with effector cells, or carry messages to a primitive network of neurons which is designed to conduct in all directions to muscle cells throughout the coelenterate body. It is well to recognize that this primitive nervous system arose from cells already in existence and probably retained metabolic systems for conduction and processes of transmission, from cell to cell, that had evolved earlier.

With the origin of bilateral symmetry in the flatworms, the earlier diffuse network began to condense and the movement of animals in a given direction resulted in the beginnings of cephalization and the alignment of cell processes into nerves. From this point on throughout the animal kingdom the basic unit plan of the reflex arc is found, but the gross anatomy of the nervous system varies with the body plan of each main animal group.

Along with the development of attenuated protoplasmic pathways as direct lines of communication another type of coördinating system evolved—the endocrine system. In this system chemical substances produced by cells in one part of the organism travel by way of circulating fluids to all parts of the body. This is a kind of 'broadcasting' system in contrast to the nervous 'telephone exchange'. Only cells which are properly 'tuned in' respond to a particular hormone. The early beginnings of the endocrine system are poorly understood, but in the arthropods such as Crustacea and insects it is found in an advanced state. The remarkable sinus glands of the crustacean eyestalks are not unlike the vertebrate pituitary in the diversity of function over which they exert control.

It is obvious in the invertebrates that a sharp distinction cannot be drawn between the modes of action of the nervous and endocrine systems, at least as regards events occurring between

[1] Atlantic City, N. J., March 1948.

cells. Neurons are designed for rapid protoplasmic conduction, the exact nature of which is still poorly understood, but at the same time they may act on an adjacent neuron or on an effector cell through the release of chemical substances. This transmission process, since it is over short distances and is fast acting, may utilize relatively unstable molecules compared with those familiar as hormones, but chemical events at cell membranes are undoubtedly common to both endocrine and nervous coördination. It has been known, since the work of Gaskell in 1914, that chromaffin cells in the ventral ganglia of the leech give rise to nerve fibers which have an excitatory action on the pulsating lateral blood vessels and that the effects of their stimulation may be reproduced by applied adrenalin. Gaskell came close to demonstrating cholinergic neurons, since he also showed that the action of inhibitory nerves on the same vessels in the leech could be mimicked by applied muscarine.

Of the chemical substances, produced by the nervous system in invertebrates and apparently involved in junctional transmission, acetylcholine is the best known. This ubiquitous substance and the enzyme systems for its synthesis and destruction were in existence long before the first differentiated nerve cell. For example, certain bacteria are effective synthesizers of acetylcholine and acetylcholine appears to be present in protozoa. The flatworms are the most primitive animals in which large and easily demonstrable amounts of acetylcholine and cholinesterase occur in nervous tissue. Acetylcholine has been found in all groups above the flatworms where sufficient isolated nervous tissue can be obtained for assay, but that we still have much to learn concerning its exact role or roles is apparent to all who follow the work in this particular field. It should be mentioned in passing that neurons in invertebrates undoubtedly produce and release neurohumors other than acetylcholine and adrenalin. However, the work on most neurosecretory cells has not progressed to the point of identifying the chemical nature of their secretory products.

I have tried to suggest that there are certain similarities between the nervous and endocrine systems in invertebrates as regards events occurring between cells. I have pointed out that cell boundaries presented certain problems in the course of evolution of the metazoa and that the need for fast, private communication simultaneously to many cells was met by certain cells becoming greatly elongated. Now by way of a more direct introduction to the subject of this symposium, a few words may be said about junctional transmission in the invertebrates. If there is difficulty in correctly evaluating the importance of chemical and electrical events at synapses and neuro-effector junctions in the vertebrates, this difficulty is greater in the invertebrates because of their variety and the infrequency with which they are carefully studied. If one follows the common practice of speaking of electrical or chemical transmission without attempting to account for the precise mode of action, or without consideration for the interdependence of electrical and chemical events in nerve and muscle, one may say that at some junctions, such as nerve to body muscle in leeches, transmission seems to be primarily a chemical event, while in arthropods it appears to be electrical. In the case of the control of heart beat in decapod Crustacea the evidence points to the involvement of acetylcholine in the action of the excitatory nerves on the heart while the inhibitory nerves appear to act without this kind of mediation. The question that now arises is—are there two strictly independent modes of junctional transmission—chemical and electrical—or are chemical and electrical events closely interwoven, with one or the other more easily demonstrated depending on the particular junction being studied or the interests and techniques of the investigator? We may hope that the papers to follow will help in answering this question.

PHYSIOLOGY OF NEURO-MUSCULAR JUNCTIONS: ELECTRICAL ASPECTS[1]

STEPHEN W. KUFFLER

Wilmer Institute, The Johns Hopkins Medical School, Baltimore, Maryland

This discussion will be confined to the 'transmitter' problem, i.e. to the events that take place between the arrival of a nerve impulse at junctions and the subsequent setting up of a muscle impulse. No attempt will be made to refer to all of the pertinent literature, particularly to earlier now quite well known papers dealing with problems of transmission.

Since the discovery of the junctional potential at the end-plate (e.p.p.) about 10 years ago (11, 16) it has been evident that this represents an intermediary process between nerve and muscle impulses. The properties of the e.p.p. have since been studied in great detail in normal and curarized muscle. It was concluded that the e.p.p. is an expression of an excitatory process, the 'transmitter', whose time course and intensity could be analyzed. From single nerve-muscle fiber preparations it became clear that the end-plate potential normally sets up the muscle impulse by depolarizing the muscle membrane around the junction to a critical level. The main problem now is not how the muscle impulse is set up, but how the nerve impulse sets up the e.p.p. As a general scheme the following sequence can be presented: nerve impulse → 'transmitter' → e.p.p. → muscle impulse. 'Transmitter' should for the present mean no more than an excitatory process, some mechanism produced by nerve impulses to bring about the e.p.p.

At this point it may well be stressed how neuromuscular transmission fits into the general picture of transmission in the nervous system. Events analogous to the e.p.p. occur at numerous cell junctions in a varied number of animals and organs. The similarity of processes may be seen from records of figure 1 at sympathetic ganglia in the cat, in the squid stellate ganglion and from potential changes in the cat or frog's spinal cord, as recorded from the ventral roots. Also the Limulus heart ganglion belongs in this class (29). All these have one common feature, namely the 'synaptic' potentials which cause the propagated impulses. 'Synaptic potential' may be defined most conveniently as the potential set up trans-synaptically by a presynaptic nerve impulse. It normally constitutes a link between two processes, having different properties from both.

The study of junctional potentials has so far been applied to crustaceans and vertebrate skeletal muscle (cat, frog) but has not been extended effectively to smooth muscle and the heart. The reason lies in the morphology of these structures which makes them at present 'inaccessible'. The local nature of junctional potential changes, which is normally swamped by the much larger propagating impulses and the shunting effect of surrounding tissue, make electrical recording techniques difficult even in isolated preparations like the sartorius of the frog. There is hope now that e.p.p.'s may eventually be recorded in human beings, particularly in some pathological conditions or after treatment with certain curare-like drugs.

The end-plate has become, in recent years, a physiological entity which at present cannot be strictly identified with the structure of the histologists, although progress in that direction is being made, particularly by Couteaux (7). Much interest is also focused on comparative studies of structures like the plates of the electric rays (14). The e.p.p., for instance, while originating in the end-plate which can be quite accurately located, involves also the surrounding muscle tissue which becomes depolarized before the muscle impulse is set up. Conversely the muscle impulse, if started by direct electrical stimulation depolarizes the end-plate while it propagates past the junction. Effective membranes separating muscle from end-plate do not appear to exist (22). There are some distinctive properties peculiar to end-plates: specific chemical excitability to substances like nicotine, caffeine and acetylcholine (ACh); this is greatly increased by denervation. No peculiar electric excitability has, however, been noted. Also K^+ does not show specific stimulating effects on junctional regions of frogs. With tetany or cal-

[1] The recent work of the author reported in this paper was aided by a grant from the National Foundation for Infantile Paralysi .

B & Z Library

cium lack the end-plate becomes involved prior to other structures. In myasthenics for instance the junctions seem to be affected first. Curare acts specifically on this area. When a nerve impulse reaches the junction it can set up a response as large as the muscle impulse but different in character; this type of response cannot at present be produced elsewhere along a muscle fiber (21).

We know practically nothing about the basis of this specialization, but from experiments on ganglion cells in changing the chemical environment (4), similar properties in other synapses, as compared with axons, may be suspected.

The physiology of a different type of end-plate from the well-known one has recently also been investigated in striated frog muscle (24). It receives small-diameter motor nerve fibers which set up a potential resembling the curarized e.p.p. of the known 'twitch system', or the prevalent local junctional potential of crustaceans. As in the latter the potential produces local non-propagated shortening of numerous muscles; it is absent in some.

The present discussion will be confined mainly to transmission in the twitch system of the frog and to the 'normal' one in mammals, the only one so far known. The analogy between these two seems complete.

The rôle of the e.p.p. in initiating muscle impulses is well shown during the progressive action of curarine (fig. 1B). The e.p.p. can be decreased by about two-thirds before neuromuscular block results and it appears that a critical minimal membrane change is required for successul transmission. This implies that the normal transmitter mechanism has a fair safety margin. Such action is, by the way, probably at the basis of the possible clinical use of this drug.

The 'transmitter' producing the e.p.p. does not cease to act after the muscle impulse has been set up but is still present for several milliseconds (msec.). It produces the subsequent potential change which is confined to the end-plate and is the first to be affected by any agent which tends to block transmission, such as curare or fatigue. While figure 2A gives a rough estimate of the persistence of the normal e.p.p. a more detailed analysis has been made by a study of the curarized potential at the junction. This is largely based on two assumptions: *a*) the potential is built up while the 'transmitter' acts and its rate of rise and magnitude is an approximate indication of 'transmitter' intensity; *b*) when the 'transmitter' has ceased, the potential declines, approximately exponentially. Such an analysis is independent of the nature of the excitatory process. In this way a total duration of about 5 msec., reaching a maximum in the first 2 msec. at 18°, has been obtained for frog's muscle (13). A different approach, while applying the same principle and giving similar results, was used in studies on interaction of nerve excitation with antidromic muscle impulses. Figure 2D serves as illustration. A nerve impulse may reach the neuro-myal junction while a muscle impulse sweeps past it, rendering it refractory. During that period no potential can be built up. As soon, however, as the membrane recovers, the surviving 'transmitter' produces a depolarization which may persist for several msec. Figure 2C shows a corresponding picture, but obtained by a different procedure. Potentials recorded at the end-plate, set up by nerve stimulation and direct muscle stimulation, were superimposed. The difference is due to 'transmitter' contribution and is confined to the end-plate region (for details see 22). The validity of such an approach is supported by experiments with applied currents or chemicals. The 'active' transmitter phase is easily reproduced by a rectangular curret pulse; if applied instead of a nerve impulse on top of an antidromic muscle spike, it produces similar membrane changes while the current flows (fig. 2B).

NATURE OF THE 'TRANSMITTER'

This much debated subject will be discussed in relation to the e.p.p. only. Since the 'transmitter' is responsible for the e.p.p. it follows that any mechanism acting on it will affect the junctional potential also. In the first place it does not seem helpful to call the cycle of transmission purely 'electrical' or 'chemical'. It has, however, become accepted to call 'electrical' the propagation of a nerve impulse in which the mechanism of spread is initiated by current flow. The impulse cycle itself is obviously complex in nature. If, in the course of an impulse a substance is 'liberated' or ions leave the membrane or if, for instance, an enzyme system is 'activated' then the term 'chemical' is generally used. This concept seems to be derived largely from hormone secretion.

'Chemical' mechanism. How then does the ACh hypothesis fit into this picture? According to this thesis ACh should be responsible for the e.p.p. and should normally disappear at a similar rate. The main support for the theory is derived from the

following evidence and based on arguments as stated in part in 1942 by Eccles, Katz and Kuffler.

ACh is liberated by motor nerve impulses; it has a powerful and rapid stimulating action on the end-plate region of the muscle; it is rapidly hydrolyzed by an enzyme concentrated at or near the end-plates. The rapidity and chemical specificity of the ACh effect suggest that ACh produces its depolarizing action by combining with specific chemical receptors on the surface. Curarine, while apparently not affecting the ACh liberation, opposes its depolarizing action on the muscle membrane. Since, in addition to this specific blocking action, curarine chemically resembles ACh (both are quaternary ammonium cations, as also are other substances with similar actions), it seems likely that it also acts by combining with these same chemical receptors (cf. the receptive substance of Langley).

If we assume that agents inhibiting choline esterase (ChE) exert their action on the ACh mechanism, and not by some other property, then the 'transmitter' effect as analyzed above, should be prolonged. This actually happens and junctional potentials lasting for seconds may be obtained. The effects are particularly striking with repeated nerve excitation. Normally the 'transmitter', although outlasting the setting up of the muscle impulse, decays before the refractoriness has recovered. With eserine, however, it may persist long enough to set up repetitive, instead of single responses.

Some difficulties arise with curare-eserine antagonism like the shortening by curarine of prolonged junctional potentials in eserinized muscle; further, for instance, even with large doses of eserine the 'transmitter' is destroyed at a fairly high rate, indicating some eserine-resistant mechanism of destruction. At crustacean junctions and in the frog's or cat's spinal cord, ACh and curarine seem ineffective although electrical events are basically similar to nerve-muscle junctions; an extension of this hypothesis to such structures is at present difficult. On the whole, however, it still offers a plausible explanation for neuro-muscular transmission. It is also conceivable that ACh would be

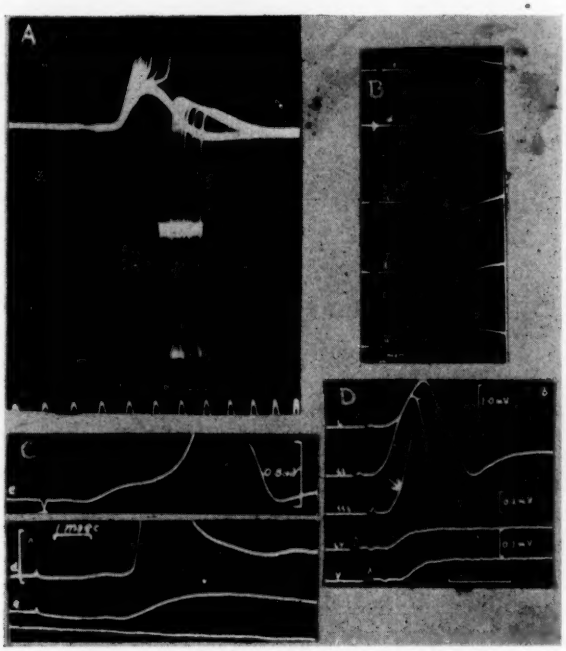


Fig. 1. POTENTIAL CHANGES at various junctions. *A*. Squid stellate ganglion, single synapse preparation, during development of fatigue. Nerve impulses (spikes off the record) arise progressively later from synaptic potential (6). *B*. Single nerve-muscle fiber preparation during progressive curarization. Muscle impulses arise progressively later from e.p.p.'s (21). *C*. Cat, potentials from ventral root in response to dorsal root stimulation. *d*. Spike from monosynaptic reflex discharge. *c* and *e*. Spike partially and fully inhibited by preceding nerve volley. Note synaptic potential (10). *D*. Cat stellate ganglion before (i, ii, iii) and after, (iv, v) complete curarization. Amplification varied. Arrow in iii shows double-step rise (8).

'liberated' or 'activated' in the end-plate region, i.e. transsynaptically, by currents generated by the nerve impulse. Since, however, an antidromic muscle impulse depolarizes the end-plate region, it would follow that it also liberated a 'transmitter'. This can certainly be excluded in eserinized muscle, where the typical prolonged end-plate potential change does not appear with antidromic stimulation.

'Electrical' mechanism. What is the rôle of current flow generated by the nerve, in producing the e.p.p.? That synaptic or nerve-muscle transmission is produced in such a way has been stated in one form or another for about 70 years and longer. Two experimental findings seem to be most pertinent in this connection: *a*) the 'electrical' nature of propagation in nerve and *b*) current flowing during a nerve impulse in one fiber can penetrate an adjacent fiber and can cause changes in threshold there.

The rôle of current spread in the propagation of nerve impulses has been demonstrated by Hodgkin (18). Part of the nerve impulse will penetrate through a blocked (pressure, cold) region for several mm. and will cause membrane depolarization and excitability changes of identical time course. Penetration of a small fraction of the nerve impulses will suffice to cause membrane changes which may bring threshold lowering of say 90 per cent. Currents will also penetrate through a cocainized region (26) and probably also through many other types of block. An essentially similar approach has been used previously by Blair and Erlanger (3), who showed that an anodal block may prevent the propagation of one nerve impulse while a subsequent one will penetrate. This observation is interpreted in the following manner:

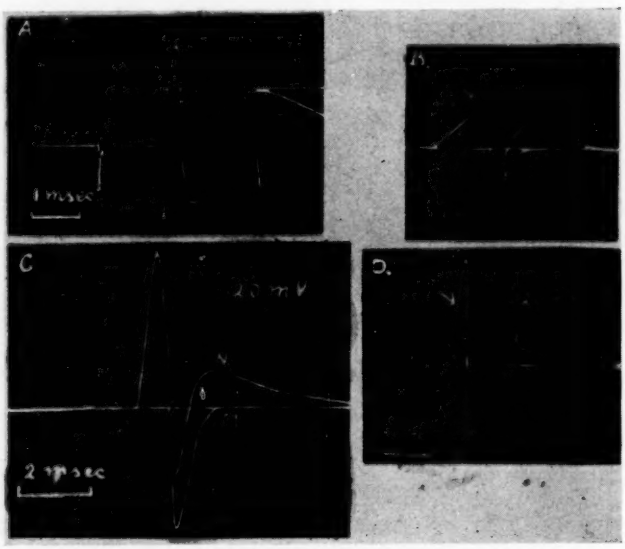


Fig. 2. E.P.P. AND 'TRANSMITTER' ACTION in single nerve-muscle fiber preparations. *A*. Interface recording, accurate end-plate position, shows late e.p.p. component following nerve stimulation. *B*. Isolated muscle fiber in paraffin oil, stimulated by prolonged current pulse, two records superimposed; *a*, current pulse withdrawn at arrow and *b*, pulse continued until diphasic wave, builds up potential. Note the absence of potential addition on top of spike; compare with *D* below *C*. Preparation in paraffin, one electrode at end-plate. *N* after nerve stimulation, *M* after direct 'antidromic' stimulation of muscle fiber. Records superimposed. *D*. Same preparation and set up; *a*, antidromic alone, *b* antidromic plus nerve impulse. Action of nerve impulse starts at arrow, but adds potential only later (22).

The analysis of Katz and Schmitt (20) and of Marazzi and Lorente de Nó (27), of the effects of penetrating currents during impulses, present impressive evidence for the efficacy of such flow from one cell to another. These and numerous other experiments on interaction of adjacent tissues during activity (1, 17, 19) greatly increase

one's confidence in the likelihood of similar events for synapses and nerve muscle junctions. A theory on these lines was recently elaborated by Eccles (9) and the principles were extended to cover also inhibition phenomena (5).

While details of Eccles' theory can be found elsewhere, a short summary of its main points follows:

1. An impulse in a pre-synaptic nerve fiber generates a current which gives a diphasic effect at the synaptic region of the post-synaptic cell, with a total duration of probably not more than 1 msec. in mammalian muscle and the spinal cord; initial anodal focus, with cathodal surround; more intense cathodal focus, with anodal surround.

2. This cathodal focus sets up a brief and intense local response at the synaptic region.

3. From this local response, a catelectrotonus spreads decrementally over the post-synaptic cell membrane.

4. A propagated impulse is set up in the post-synaptic cell, if this catelectrotonus is above a critical value. If it is below, then, as the local response subsides, the catelectrotonic surround decays passively.

One of the new contributions of Eccles in this scheme was the insertion of the local response, thereby obtaining a mechanism for setting up the large post-synaptic response. This would enable relatively weak currents to effect transmission.

In an attempt to test the stimulating efficiency of currents penetrating from the nerve terminals to the end-plate, a few preliminary experiments have recently been completed. They were largely prompted by Eccles' new challenging theory and by numerous exchanges of ideas with Dr. B. Katz of University College, London. *a*) Nerve impulses were blocked by an anode within a mm. of the nerve terminals while observing the potential changes at the end-plates. *b*) Subthreshold cathodal shocks of different strength and duration were applied to the terminals while recording nearby at the junction. *c*) Neuro-muscular delay times were measured after the nerve had been stimulated near the junctions. All experiments were done on dissected preparations, containing about 3 to 4 muscle fibers with their intact nerve supply. The semitendinosus muscle was preferred for these dissections. Under the microscope nerve branches innervating adjoining fibers within a small area could easily be located. It was essential to clean the nerve fibers right down to the junctions. In single nerve-muscle fiber preparations this is more difficult than in such small bundles and furthermore the nerve can be handled in the latter more conveniently without injury, particularly by pulling. It was thus easy to place stimulating electrodes (100 μ diameter) at the nerve entry. The next essential feature is accurate recording from the adjoining end-plates. In single nerve-muscle fiber preparations the initial deflection resulting after nerve excitation may be entirely composed of the e.p.p. component (21). Such a recording position then can be regarded as covering the entire end-plate region. In the present experiments only those preparations were used where an initial e.p.p. component of about 50 per cent of spike height was obtained and where a shift of 100 micra caused a marked fall of that component. Such end-plate localization was always within 0.5 mm. of the dissected nerve entry. Scatter of end-plates is easily detected. The preparation was either completely surrounded by paraffin oil or the muscle fibers were kept at the oil-paraffin interface, with the nerve penetrating the interface into the paraffin (21). While both types of recording give the same results, the latter method was preferred since more stable and accurate recording conditions were obtained.

Blocking of nerve impulses. Chlorided fine silver electrodes for blocking conduction could be placed conveniently at any portion of the nerve, entirely surrounded by paraffin oil, while the nerve trunk was stimulated above. Since the muscle fibers were generally innervated by not more than 2 to 3 nerve fibers, block of conduction usually occurred simultaneously in all. The block, however, could be kept just critical as determined by the following test. (*a*) When stimulating at a frequency around 3 per second occasionally a twitch of the muscle fibers would result, indicating penetration past the block. (*b*) A second impulse following 10 to 50 msec. after the first one would regularly set up normal muscle activity, thus 'jumping' across the block (3).

Nerve impulses blocked in such a manner did not exert an appreciable potential change on the nearby end-plate region. A change of 1 to 5 per cent of the normal initial e.p.p. component would have been easily detected. The stimulating efficiency of penetrating currents, under similar conditions, extending several mm. past an anodal block, has been fully demonstrated in single nerve fibers by Blair and Erlanger.

Subthreshold stimulation of nerve terminals. At the cathode of an applied current pulse an extrapolar spatial and temporal spread of electrotonic

potential results. If a pulse of near-threshold strength is applied to the nerve portion within 1 mm. of the end-plate, then the terminal region should become partially depolarized. This in turn will cause current flow from the non-affected parts into this area, part of it penetrating the nearby end-plate region. The experiment is essentially similar to the previous one, where the blocked impulse furnished the subthreshold excitation. Beside brief shocks, double condenser discharges of varying shape and duration were used. Such currents passing through the junctional region did not cause any depolarization exceeding several per cent of the normal e.p.p. size.

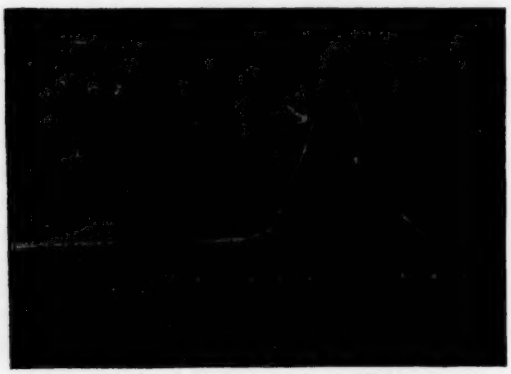


Fig. 3. SEMITENDINOSUS, ISOLATED PREPARATION of 3 to 4 muscle fibers and nerve supply. Interface recording at 22.5°C. Stimulating cathode within 0.5 mm. of end-plate electrode. Stimulus 3 times threshold. Stronger stimulus does not reduce shock-response period. Arrow marks initial e.p.p. size. Time, 5000 c.p.s.

Neuromuscular delay. Neuromuscular delay and its analogue, synaptic delay, may be defined as the time between arrival of a nerve impulse front and the start of the e.p.p. or synaptic potential. Synaptic delays for mammals have been calculated for the central nervous system by Lorente de Nó, Eccles, Lloyd and others. There seems to be agreement for values ranging between 0.5–0.9 msec. In those values no account is taken of synaptic potentials. The neuromuscular delay for frog sartorius has been calculated as between 1.1 to 1.3 msec. at 20° (13). A somewhat more accurate estimate can be obtained by stimulating the terminal nerve region while recording from the end-plates. Such a record is illustrated in figure 3. A maximal cathodal shock is followed after 1.0 msec. by the rise of the initial e.p.p. Further strengthening did not shorten the shock-response period. This time then may be assumed to consist of conduction time, 'setting-up' time and the true delay period.

It seems reasonable to assume that the nerve fibers were excited within half a mm. of the actual end-plate. Nerve conduction time over this stretch would be negligible at speeds prevailing in the nerve trunk. It was found that conduction velocities of the motor nerve fibers of this preparation were around 30 m/sec. at 22°C. Besides the delay itself there remains then the time for 'setting up' a nerve impulse. With condenser discharges (also used in fig. 3), rising sharply to a peak in 0.05 msec. and falling to one-half in a further 0.1 msec., the earliest nerve impulses recorded near the stimulating cathode in isolated nerves arose in 0.1 msec. after onset of the condenser discharge. With shocks nearer threshold the latent period increased. Under these conditions, unless the nature of electrotonic spread is radically different over the 0.5 mm. stretch before the end-plate, one may assume that the terminal region is greatly affected practically instantaneously by electrotonic spread and subsequently by the 'active' portion of the spike, probably within the first 0.2 msec. (cf. later). There remains then a period of about 0.8 msec. delay in the preparation of figure 3. In other preparations the times calculated on this basis varied between 0.5 and 1 msec. at about 22°C. During the delay no intermediary potential changes were detected at the end-plate preceding the e.p.p. This was ascertained several times by crushing the nerve terminal, causing block. The subsequent record did not differ in its course up to the initial e.p.p. deflexion and was nearly identical with the normal base line at just subthreshold stimulus strength.

If the nerve terminal is radically depolarized during the greatest part of the delay period, current flow into that area should occur from the surround. Further, if we assume that the e.p.p. is set up by the penetrating currents from the terminals then we must account for the lag between the two processes. Eccles believes that factors like the initial anodal focus and the time constant of the transsynaptic or transjunctional region, delaying the building up of the potential could be an adequate explanation. With stimulation practically at the end-plate, however, the duration of the initial anodal current should be extremely short probably of the order of 0.2 msec., if that is the time assumed for the active part of the nerve impulse to reach the terminals with strong stimuli; it is uncertain therefore whether relatively

long delays (cf. also later) can be explained on such a basis.

In order to elucidate events in terms of current distribution, it was attempted to reconstruct lines of current flow for the present situation similar to the scheme of Eccles. The origin of the cathodal effect at the junction is different if an impulse is set up at the end-plates and propagates away, or when the impulse approaches. Figure 4 illustrates the scheme of flow when the nerve impulse is set up within a mm. of the junction. There would be a short anodal focus and cathodal surround created transjunctionally lasting until the active part of the impulse reaches the endings (fig. 4, 1). The subsequent situation is more difficult to visualize and is certainly different from the usual one, when the impulse approaches the end-plate (9). It is drawn in two sections. While the impulse spreads to the terminals it also extends in the opposite direction, creating a stretch of about 4 to 6 mm. (at 20–30 m/sec. conduction velocity) of nearly uniformly depolarized nerve. The cathodal focus at the end-plate would then be provided by currents generated by the advancing impulse drawn in figure 4, 2a. This cathodal contribution would diminish as the impulse travels further away from the end-plate. During the same period the site where the impulse was initiated would start to recover and a field of current distribution as seen in figure 4, 2b, would develop and provide the cathodal focus until the terminal nerve region has been largely restored, probably within 1 msec. By that time the impulse front leaving the junction would be at least 20 mm. away, thus having no or negligible effect.

It remains questionable whether such a scheme provides an adequate cathodal focus. In this connection it is interesting, however, to note that the e.p.p. shape set up with both types of excitation (nerve impulse approaching or leaving end-plate region) is not appreciably different.

Duration of penetrating current and of 'transmitter' effect. Experiments on the interaction of nerve impulses and antidromic muscle impulses (cf. above) provide an additional difficulty for an 'electrical' scheme. If a nerve impulse reaches the neuromuscular junction during the rising phase of a muscle impulse, set up antidromically, then it does not produce an added potential change until the refractoriness has partially disappeared. This period may be 0.5 msec. (fig. 2D) but can also be nearly twice as long. After that delay the 'transmitter' proceeds to affect the membrane, building up the later portion of the e.p.p. (Incidentally, by this type of experiment the effective neuromuscular delay—arrival of nerve impulse to start of rise of e.p.p. (cf. above)—may be prolonged by a further msec.) According to similar analyses then the building up process underlying the e.p.p. may last about 5 msec. in frog muscle (22). In terms of current flow this implies that the nerve impulse in the terminals produces transjunctionally penetrating currents of a similar duration instead of the generally assumed short phase of about 1 msec. calculated for axons (9, 20, 27).

In view of such difficulties one may postulate (9) that normally the short penetrating current sets up a local response which has a time course

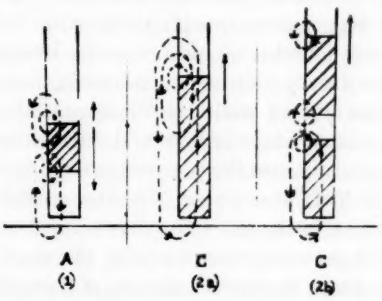


Fig. 4. DIAGRAM OF CURRENT FLOW at myoneural junction. 1. Impulse set up at nerve within 0.5 mm. of end-plate. 'Active' part of impulse shown by shaded area. Brief anodal focus *A* and cathodal surround created at end-plate. *2a* and *2b*. Nerve impulse has reached terminals, creating cathodal focus at *C* end-plates. Actual current distribution should be combination of two diagrams (see text).

similar to the subsequent muscle potential. This assumption, however, encounters difficulties in view of the interaction experiments (see *above*). The penetrating currents would have to set up the local response in 'completely' refractory muscle which does not permit even a small additional depolarization since the end-plate area is already depolarized by the antidromic impulse. Surviving that period the local response would 'grow' and then decay after several msec. It is questionable whether a local response can be set up or maintained under such conditions. Alternately, the duration of the penetrating currents has to extend over the entire 'active' phase of the transmitter action.

GENERAL DUSCUSSION

The preliminary experiments reported here in connection with 'transmitter' problems represent an attempt at elucidating the rôle of action currents in transmission. They are mainly based on

the assumption that penetrating currents should produce an electrically detectable membrane change. If, then, the e.p.p. is set up by current flow generated by the nerve impulse, the junctional region should be affected by certain changes in the nerve terminals provided that currents can establish an adequate circuit between endings and end-plates. In the case of subthreshold excitation of the terminal nerve regions without subsequent transjunctional potentials, it could possibly be argued for instance that the effective penetrating component was limited by the lack of 'breakdown' and consequent lowered resistance pathway. Such an argument does not hold if the nerve is maximally stimulated near the end-plate. The relatively long neuromuscular delay with the end-plate not affected in spite of such (intense) excitation nearby may be used as an argument for a discontinuity of excitatory mechanisms. However, one may also attribute to the nerve terminals very special properties preventing a quick current spread. This also would give the anodal focus more time to develop. The nerve impulse in the same fibers would travel during the usual delay period about 30 mm. If a slowing of the excitation spread in the terminal region by a factor of 30 to 50 as compared with the other parts of the axon did occur, one may attempt to account for the setting up of the normal e.p.p. by current spread. The quick tapering off of the final branches makes a greatly slowed conduction plausible. One must not assume, however, that, in order to set up a postjunctional potential, the impulse has first to reach all the actual terminal points. Since the fine nerve branches lie in apposition to the end-plate over a definite distance before terminating (7), one should expect the potential change to start as soon as the impulse enters that last stretch. A correlation between conduction velocities in nerve fibers and the delay period at junctions is of interest in this connection. It can be done accurately for nerve-muscle junctions. A sufficient number of experiments is not yet available.

The observations reported here were, in fact, undertaken in the belief that supporting evidence could be obtained for the effective transsynaptic action of current spread. They alone do not disprove that type of transmission (cf., however, below). In face of evidence demonstrating the rôle of current flow in conduction and also in producing excitability changes in adjacent nerve fibers, a similar mechanism acting across junctions but especially central synapses remains a plausible assumption. It naturally need not be assumed that the same mode of transmission pertains to all nervous tissues. The configuration of the tissue junctions, like assymmetry of contact, also plays an evident rôle in schemes of transmission particularly in case of current spread. The absence of facilitation and the 1:1 ratio between incoming nerve impulses and transjunctional discharges in the normal 'twitch' system is in contrast to the general picture in the central nervous system. A similar mechanism of facilitation is, however, evident in the small-nerve motor system of skeletal muscle or in crustacean nerve-muscle transmission. The electrical scheme is a proposition of current density which may not be achieved in the relatively large neuromuscular junction, for instance owing to fine branching of the nerve terminals, and this may be the reason why predictions based on observations on current spread elsewhere do not come true in myoneural transmission. A sufficient density, however, may result at small synaptic knobs.

Other difficulties of an 'electrical' explanation have been pointed out by Eccles (9). The curare-block remains unexplained, the effect of anticholinesterases on the e.p.p. points to a rôle of an ACh-like transmitter. Long delays in some systems have to be considered (15, 30). An assumption of the duality of transmitters supplementing each other may be, to many, an obvious, but unattractive solution of difficulties.

The inhibitory nerve impulse in crustaceans reaching the end-plate region and acting on the membrane and on the linkages beyond it, i.e. by an intracellular mechanism, without changing its potential, is a further difficulty pointing out the possible ineffectiveness of mere current spread from a nerve impulse (25, 28). It emphasizes the well known fact that membranes can be greatly affected without appreciable change in resting potentials. Curarine at junctions and novocaine or cocaine in nerve or muscle being just two examples (2). There may be similar mechanisms operating in the physiological processes of the vertebrate nervous system. They are, however, rarely considered to be part of inhibitory function, i.e. nerve impulses purely inhibitory in nature, conditioning the membrane but not changing its potential. Conversely, also, sight has been lost of Gaskell's (1886) old, but several times confirmed observation, that the inhibitory action of vagus stimulation is accompanied by an increased polarization of the heart muscle. An adequate histological picture of crustaceans, showing the

end-plates receiving both types of nerve fibers, is of primary importance.

Difficulties from experiments with veratrine. A further serious difficulty of effective action current penetration into the transjunctional region arises from experiments with veratrine. These have not been discussed before in this connection. It should be an essential feature of any 'electrical' theory that the nerve impulse generates the current flow setting up the e.p.p. and, consequently, prolonged potentials in the nerve terminals should also be accompanied by prolonged transsynaptic changes. Such experimental conditions are obtained by veratrine treatment of preparations. Veratrine increases the normal after-potential in axons and certainly should do the same in the terminals. No concommitant prolongation of e.p.p.'s, however, could be observed in well controlled experiments (23). Such tests should also be applied to synapses like ganglia. This type of experiment also appears to exclude the afterpotential from a primary rôle in normal transmission. It seems significant that, in line with the present argument, veratrine does not change appreciably the short spike component of an impulse. Production of 'transmitter', therefore, should be associated with this part of the nerve impulse only, probably with the period of 'breakdown', which is unlikely to be much longer than 1 msec. To account for the analyzed 'transmitter' duration, the breakdown would have to last around 5 msec. in the terminals.

It appears from the present experiments that for successful transmission the nerve terminals have to be completely depolarized, since intermediary activation does not produce any graded transjunctional changes (cf. also afterpotentials). It is conceivable that permeability changes of the nerve terminals would affect the end-plate membrane lying in close apposition. Although intimate structural details are not known it may be assumed that many slowly moving ions could reach the transjunctional region even during quite a short delay period. No evidence for any intermediary tissue between terminals and muscle surface exists (however, cf. 7). Theoretically, ions like K^+ could set up the e.p.p.'s, but until good evidence for some other mechanism becomes available it seems more reasonable to assume that during the breakdown some stored ACh-like substance is released in the terminals.

SUMMARY

The discussion is confined to the 'transmitter' problem, i.e. to the mechanism by which the nerve impulse gives rise to the end-plate potential. Duration and intensity of the 'transmitter' is analyzed. In this connection the ACh and action current theory of transmission is briefly discussed. Some new experiments, testing the efficacy of current flow in the terminal nerve region on the transjunctional region, are presented. Potential changes were recorded from the end-plate regions of isolated preparations, while stimuli were applied within 0.5 mm. to the nerve nearby. Subthreshold depolarization of the nerve terminals by applied currents or by blocked nerve impulses did not affect appreciably the end-plate membrane. Neuromuscular delay periods were measured and are discussed in relation to current spread from nerve to muscle.

It is concluded that *a*) action currents in the terminals are not effective in depolarizing the end-plate region. *b*) The 'transmitter' action occurs during the 'breakdown' in the terminals. The 'breakdown' period is too short to account for the analyzed 'transmitter' duration. Prolongation of the depolarization in the terminals does not prolong transjunctional potentials. It is thought that ions liberated during the 'breakdown' in the nerve terminals could best account for the observed phenomena.

ADDENDUM

Since this symposium was held a communication from Dr. Eccles has been received, stating that the electrical hypothesis cannot be reconciled with more recent experimental results on neuromuscular transmission. Eccles and his coworkers now believe that their evidence favors ACh as the sole mechanism (*Ann. Rev. Physiol.*, 1948).

REFERENCES

1. Arvanitaki, A. *J. Neurophysiol.* 5: 89, 1942.
2. Bishop, G. H. *J. Cell. and Comp. Physiol.* 1: 177, 1932.
3. Blair, E. A. and J. Erlanger. *Am. J. Physiol.* 126: 97, 1936.
4. Bronk, D. W. *J. Neurophysiol.* 2: 380, 1939.
5. Brooks, C. McC. and J. C. Eccles. *Nature* 159: 760, 1947.
6. Bullock, T. H. *J. Neurophysiol.* 11: 343, 1948.
7. Couteaux, R. *Rev. Canad. Biol.* 6: 563, 1947.
8. Eccles, J. C. *J. Physiol.* 101: 465, 1943.

9. Eccles, J. C. *Ann. N. Y. Acad. Sci.* 47: 429, 1946(a).
10. Eccles, J. C. *J. Neurophysiol.* 9: 87, 1946(b).
11. Eccles, J. C. and W. J. O'Connor. *J. Physiol.* 97: 44, 1939.
12. Eccles, J. C., B. Katz and S. W. Kuffler. *J. Neurophysiol.* 5: 211, 1942.
13. Eccles, J. C., B. Katz and S. W. Kuffler. *J. Neurophysiol.*, 4: 362, 1941.
14. Fessard, A. *Ann. N. Y. Acad. Sci.* 47: 501, 1946.
15. Gilson, A. *Am. J. Physiol.* 138: 113, 1942.
16. Goepfert, H. and H. Schaefer. *Pflüg. Arch. Ges. Physiol.* 239: 597, 1937.
17. Granit, R. and C. R. Skoglund. *J. Physiol.*, 103: 435, 1945.
18. Hodgkin, A. L. *J. Physiol.* 9 : 183, 1937.
19. Jasper, H. H. and A. M. Monnier. *J. Cell. Comp. Physiol.* 11: 259, 1938.
20. Katz, B. and O. H. Schmitt. *J. Physiol.* 97: 471, 1939.
21. Kuffler, S. W. *J. Neurophysiol* 5: 18, 1942(a).
22. Kuffler, S. W. *J. Neurophysiol.* 5: 309, 1942(b).
23. Kuffler, S. W. *J. Neurophysiol.* 8: 113, 1945.
24. Kuffler, S. W. and R. W. Gerard. *J. Neurophysiol.*, 10: 383, 1947.
25. Kuffler, S. W. and B. Katz. *J. Neurophysiol.*, 9: 337, 1946.
26. Lorente de Nó, R. *J. Neurophysiol.* 2: 402, 1939.
27. Marazzi, A. S. and R. Lorente de Nó. *J. Neurophysiol.* 7: 83, 1944.
28. Marmont, G. and C. A. G. Wiersma. *J. Physiol.* 93: 173, 1938.
29. Prosser, C. L. *J. Cell. Comp. Physiol.* 21: 295, 1943.
30. Rosenblueth, A. and R. S. Morison. *Am. J. Physiol.* 120: 384, 1937.

PHYSIOLOGY OF NEURO-MUSCULAR JUNCTIONS: CHEMICAL ASPECTS

GEORGE H. ACHESON

From the Harvard Medical School, Boston, Massachusetts

Since the beginning of the war, the chemical aspects of neuromuscular transmission have advanced at a slow pace. Little that can be recognized as fundamental has been discovered. A number of subjects connected with this field have undergone considerable development; but for the most part this development is in directions which do not contribute to an understanding of the neuromuscular synapse itself. The most significant advances in neuromuscular transmission have been in the electrical realm. The new electrical information leads us to a new perspective, enabling us to re-evaluate some of the older observations, and by a combination of old and new to push forward the frontier.

MICROSCOPIC ANATOMY

Since the new physiological data have brought into new perspective the minute structure involved in transmission, it is particularly important now to have a clear notion of the morphology of the neuromuscular synapse. The microscopic anatomy of this region has advanced significantly.

For 90 years it has been known that the motor nerves to a skeletal muscle fiber end in an arborization at a special region of the muscle. Much attention has been given to the forms taken by the nerve endings. They have usually been examined after impregnation with silver or gold. The work of Couteaux (1) has been based on another histological method, the supravital use of the dye, Janus green. It focuses our attention on a structure which is separate from and outside of the terminal arborization of the nerve. Material which stains with Janus green is arranged in palisades. Rod-like masses at the limit of optical resolution appear to be lined up side by side perpendicular to the surface of the nerve terminals. From the ends of these which face the muscle, fine, thread-like projections extend into the surrounding sarcoplasm. Cross sections of end-plates show that the nerve endings lie in gutters or grooves. The material which stains with Janus green lines these gutters on the muscle side.

This arrangement of the Janus green-staining material has been found in the mouse and guinea pig and in two species of lizard. In the frog, it is less well developed. Palisades are arranged along the course of one of the fine nerve endings which constitute the branches of the end bush described by Kühne as characteristic for the frog.

These unusual features of the frog's neuromuscular synapse remind us of the importance of comparative observations in different species and different muscles, and in other organs of similar behavior. In the case of the electric organs, the evidence indicates that the single unit, the electroplax, is like a neuromuscular end-plate, without

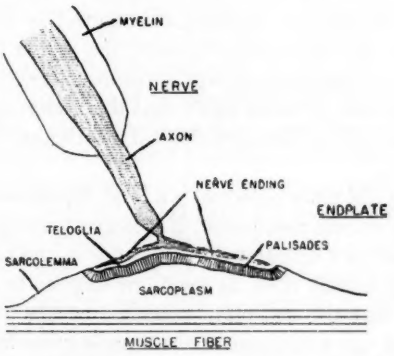


Fig. 1. DIAGRAM OF END-PLATE REGION, adapted from Couteaux.

the accompanying muscle fiber, with its conductile and contractile systems. It has long been recognized that the electroplax contains palisades, lining the region where the terminal arborization of the nerve makes contact with the electroplax. The resemblance between these palisades and those described by Couteaux in the end-plate is striking. It supports the hypothesis that this material is of some physiological importance both in the electroplax and in the end-plate.

The diagram shown in figure 1, adapted from Couteaux, summarizes our knowledge of the morphology of the end-plate region. Couteaux believes that a special sheath, the teloglia, with its own nuclei closely inverts the nerve terminals.

CHEMICAL SPECIALIZATION OF THE END-PLATE REGION

The events of neuromuscular transmission represent physiological activity of the structures indicated in figure 1. The nerve impulse arrives above and the muscle impulse departs to either side. Some structure in this region is responsible for the end-plate potential. Is it the teloglia? Or the palisades? Or is it some structure as yet unknown?

The nerve has a physiological membrane across which a potential is maintained by oxidative metabolism. Likewise the muscle has a membrane with a membrane potential. The exterior of these two physiological membranes face each other in the region represented by figure 1. The arriving nerve impulse and the departing muscle impulse represent waves of depolarization of the respective membranes. When the sarcolemma is pierced anywhere along the muscle fiber by a microelectrode, the potentials which are recorded indicate that the tip of the electrode is inside the place where the membrane potential is maintained (2). Since histologically, the region lined by the palisades appears to be continuous with the sarcolemma, it seems likely that this region represents a part of the physiological membrane of the muscle.

The end-plate potential is a local depolarization of the muscle membrane. If it is great enough, it can set up a conducted disturbance in the neighboring muscle fiber. It seems reasonable to suppose that the end-plate potential occurs in the part of the membrane of the muscle immediately adjacent to the nerve endings, that is, the very region which is lined with these palisades. What, then, do these palisades contribute to the behavior of this region which makes it different from other regions of the physiological membrane of the muscle and nerve?

There is excellent evidence for a specialized function of the end-plate region. This is the part of the muscle fiber where, as Langley demonstrated (3), the stimulating action of nicotine occurs and curare prevents this stimulating action. He proposed that there exists at the end-plate a specialized material, which he designated the 'receptive substance.' In recent years it has been shown in single muscle fibers that the end-plate region is depolarized by very small concentrations of acetylcholine or nicotine or by somewhat larger amounts of caffeine (4, 5). The rest of the surface of the muscle is a thousand or more times less sensitive to these substances. The depolarization which they produce in the end-plate region of the muscle membrane elicits conducted responses in the muscle fiber. Thus, as a result of the great sensitivity of the end-plate to these substances, they are able to stimulate the muscle fiber in very low concentrations. It seems likely that many other substances which possess a stimulating action on skeletal muscle similar to that of acetylcholine or nicotine also act by depolarizing this specialized part of the muscle membrane.

Langley demonstrated an antagonism between nicotine and curare. A similar antagonism exists between curare and acetylcholine. Curare is able to prevent the depolarization of the end-plate region by these substances and hence to prevent their stimulating action. The available evidence suggests that curare should be considered a competitive inhibitor of the reaction between acetylcholine or other substance and the receptive substance (6). Besides the group of alkaloids collectively known as curare, a number of other substances are capable of inhibiting the action of acetylcholine and its congeners on skeletal muscle. Many quaternary ammonium cations do this. For example, the simplest of these, tetramethylammonium ion, in certain concentrations is capable of inhibiting the action of acetylcholine at skeletal muscle. But under other circumstances this substance has stimulating properties resembling those of acetylcholine. The same is true of a large number of other quaternary ammonium compounds as well as the tertiary amine, nicotine. It seems likely that all these substances act at the specialized part of the muscle which we recognize as the end-plate. And it also seems likely that they produce their actions either by depolarizing the end-plate, or by preventing its depolarization.

The actions of potassium on skeletal muscle are in some ways similar to those of acetylcholine. Potassium, too, can depolarize the end-plate. But in this instance, the ability to depolarize is not restricted to this region. Potassium depolarizes the muscle to about the same extent in all regions. Significantly, curare does not prevent the depolarization, or the consequent stimulation, produced by potassium. Magnesium makes a block superficially similar to that of curare, both at the neuromuscular synapse and in autonomic ganglia (7, 8). But magnesium differs from curare by blocking not only the stimulating effect of

acetylcholine, but also that of potassium. Calcium also has significant effects on the end-plate region, which are to some degree separable from its effects on the membranes of nerve and the rest of the muscle fiber (9, 10).

The most striking fact about the substances mentioned which specially affect the neuromuscular synapse is that all of them are cations. Of the organic compounds, only nicotine is not a strong base. It is possible that even with nicotine, the cationic form is the effective one. The end-plate region is probably merely a specialized part of the physiological membrane. The membranes of both nerve and muscle are importantly affected by cations. Sodium is not important to the maintenance of the membrane potential, but it is essential to the mechanisms underlying excitation, and conduction, and the action potential. Potassium in excess of that ordinarily found in the extracellular phase depolarizes the membrane; it is moreover, in some way pumped into the interior of these cells, presumably by a mechanism residing in the physiological or the anatomical membranes. Calcium and magnesium have significant effects on the behavior of the excitation and conduction systems of muscle and nerve which depend on the properties of the physiological membrane.

The specialization of the end-plate membrane consists in the fact that it reacts with a group of cations which do not affect the rest of the membrane of muscle and the membrane of nerve. There may, of course, be additional ways in which the end-plate is specialized.

What accounts for the specialization of the end-plate in relation to cations? Two hypotheses have been proposed. It has been suggested that the end-plate lacks a protective covering which is present elsewhere. In particular it was proposed that the myelin of nerve prevents quaternary ammonium ions from reaching the essential parts of the conducting system and hence prevents an action of these substances on conduction (11). The actions of quaternary ammonium cations at the end-plate were supposed to be the same as those which might occur at the surface of the nerve if the myelin were not present. This hypothesis has no experimental support.

The other hypothesis to account for the specialization of the end-plate region is that there is a chemically differentiated substance at this region which is absent elsewhere in the membranes of muscle and nerve. We can designate such a substance as the receptive substance, after Langley. It is at present defined only in terms of its reactions in intact cells. It is therefore highly hypothetical and so it will remain until further data are obtained.

What parts of the structure of the neuromuscular synapse give it its specialization? Are the palisades the receptive substance? Or do they represent the large amount of cholinesterase which has been demonstrated to be concentrated in this region? Or finally, are they associated with some accessory physiological processes having to do with the end-plate potential?

THE TRANSMITTERS

What is the nature of the physiological processes of neuromuscular transmission? The new knowledge of the end-plate mechanism makes it clear that we must consider two processes: the transmission of the influence of the nerve impulse to the end-plate mechanism and the transmission of the influence of the end-plate response to the conduction system of the muscle.

At present, the second type of transmission is the easier problem to deal with. So similar is the behavior of this transmission to that of the electrical excitation of muscle fibers that the identity of the two phenomena has been generally assumed. It is proposed that the transmitter of the end-plate response is the electric current set up in the nearby muscle fiber by the local depolarization at the end-plate. The end-plate is considered to be a region specially arranged to produce this depolarization and thus to initiate muscle impulses.

The transmission from nerve to end-plate is more difficult to deal with. Recent years have brought remarkable advances in the understanding of the electrical forces involved in conduction along the nerve axon. In a number of instances, the excitatory influence of biological potentials across gaps between electrically excitable tissues has been clearly established. As a result there has been an increasing tendency to consider the transmitter of the influence of the nerve to the end-plate as the electric current produced by the arrival of the action potential at the nerve ending. These developments have tended to crowd from our minds the data gathered a few years earlier which seemed to establish that the neuromuscular transmitter was acetylcholine. It seems worthwhile to review the evidence for the two hypothetical transmitters, the electric current, on the one hand, and acetylcholine, on the other, in the

light of the present knowledge of the events of neuromuscular transmission.

In the first place it is agreed that the arrival of nerve impulses at the terminal arborization is attended by the production of electric current and by the release of acetylcholine from the nerve endings.

It is further agreed that acetylcholine can do to the end-plate mechanism what the arrival of nerve impulses can do. That is, acetylcholine can depolarize the end-plate and thus set up muscle impulses. It has been assumed that electric current (similar in duration to the action current of the nerve endings) can set up end-plate potentials similar to those elicited by nerve impulses. Direct evidence for this is, however, lacking. And a number of attempts to demonstrate such a fact have been negative.

In the absence of direct evidence that electric currents of the duration of action potentials of motor nerve fibers are capable of setting up end-plate potentials, we may well ask: Is the end-plate mechanism electrically excitable at all? To some it may appear unlikely that a structure which produces an electrical response should be electrically inexcitable. However, the available evidence suggests that electric organs are electrically inexcitable. Whenever electric stimuli elicit the electric discharge, the latter occurs at a delay which indicates that the stimuli were acting on the nerves rather than the electroplaxes. When the electric organ is chronically denervated, it becomes electrically inexcitable (12). The similarity of the electroplax to the neuromuscular end-plate is indicated by embryological, histological, physiological and biochemical data. The evidence that the electric organ is electrically inexcitable is, like that for the end-plate, negative evidence. Certainly the hypothesis that the transmitter from nerve to end-plate is electrical would be on firmer gound if it could be demonstrated that the end-plate mechanism is electrically excitable.

Are the two hypothetical transmitters produced in great enough quantity to elicit the end-plate potential? Our data are scanty. Since we do not know the electrical excitability of the end-plate mechanism, we cannot assess the adequacy of the current produced by the action potential of the nerve.

Work with single fibers has provided evidence as to how much acetylcholine is necessary to set off muscle impulses from the end-plate region. In the lizard, 5 micro-micrograms of acetylcholine applied to the end-plate, in a droplet of about 50 times the volume of the end-plate, was the minimal effective dose (4). Since at best less than one quarter of the surface of the droplet could have been in contact with the end-plate, probably, only a fraction of the dose actually diffused through the sheaths surrounding the end-plate to the receptive substance. That this fraction represented an excessive dose is indicated by the fact that the muscle responded with a brief tetanus and the end-plate was thereafter inexcitable to nerve impulses until it had been washed. We can at present do no better than to say that a fraction of 5 micro-micrograms of acetylcholine is needed to set up muscle impulses at the end-plate of a single muscle fiber.

No data are available on the quantity of acetylcholine released at single end-plates. It has been necessary to work on multifibered preparations and no direct counts of the number of nerve fibers or nerve endings involved have been made. Using data culled from the literature, I have calculated the output of acetylcholine per nerve impulse per end-plate from the gross outputs recorded by 5 groups of investigators in the literature (table 1). The quantity is 10^{-10} μg. This value is only about 1/30,000 of the minimal amount of acetylcholine necessary to stimulate muscle at the end-plate. All of the experimental errors would tend to decrease the gap. But even if the minimal dose necessary is ten times too high and the output of the nerve is ten times too low, a considerable gap remains between the two values. We are little better able to account for the effect of the nerve impulse in terms of the output of acetylcholine than we were 10 years ago when a gap of similar width was found on the basis of the minimal effective dose given by quick intra-arterial injection.

Is each of the two hypothetical mediators produced and removed quickly enough to account for the electrical events of transmission? The rapidity of appearance of the action potential is adequate; but its influence is probably too brief to account for the long synaptic delays which occur. We have no direct data on the rapidity of the release of acetylcholine. If the end-plate potential is a response to acetylcholine released by the nerve, the time course of the release must be less than the synaptic delay. With the reasonable assumption that the effect of the cholinesterase inhibitors on the end-plate potential is uniquely a protection of acetylcholine from hydrolysis, we can deduce that acetylcholine is being hydrolyzed already

when the end-plate potential is reaching its peak. The decrement of the end-plate potential indicates that if acetylcholine is the mediator, its disappearance must be almost as rapid as its release. Within these limits the acetylcholine might be present considerably longer than the action potential.

fault as indicated on the left in table 2. First, the transmitter of the influence of the nerve to the end-plate may be inadequate. In this case the end-plate potential is small and the muscle fiber fails to respond. Second, in the presence of normal delivery of transmitter, the end-plate may be unable to respond adequately. The muscle is

TABLE 1. CALCULATION OF OUTPUT OF ACETYLCHOLINE PER NERVE IMPULSE FOR EACH NERVE END [1]

	1 ACETYLCHOLINE RELEASED PER VOLLEY	2 ESTIMATED NUMBER OF FIBERS STIMULATED	3 INNERVATION RATIO	4 ESTIMATED NUMBER OF ENDINGS	5 ACETYLCHOLINE RELEASED PER NERVE ENDING PER IMPULSE
Perfused organs					
A. Cat tongue (Dale, etc.)	10^{-4} γ	2400 (A)	1:150 (D)	720,000 (E)	1.4×10^{-10} γ/Ep
B. Cat ganglion (Feldberg, etc.)	10^{-4} γ	3900 (B)	1:32 (B)	1,000,000	10^{-10} γ/Syn
Cut end of nerve trunk					
C. Frog sciatic (Lissak) (Scheinfinkel)	2×10^{-7} γ	1000 (C)			2×10^{-10} γ/Fiber
			MEAN DIAMETER OF FIBERS	ESTIMATED SURFACE OF NERVE IN ENDPLATE	
Quick-frozen nerve trunk					
D. Frog sciatic (Muralt)	10^{-8} γ per μ length	1000 (C)	10 μ (C)	500 μ^2	1.5×10^{-10} γ/Ep

A. Langworthy.
B. Billingsley and Ranson.
C. Gasser and Erlanger.
D. Clark.
E. Katz and Kuffler.

$$\frac{\text{Quantity produced}}{\text{Quantity required*}} = \frac{1.5 \times 10^{-10}\gamma/\text{Ep}}{5 \times 10^{-6}\gamma/\text{Ep}} = \frac{1}{30{,}000}$$

* (Buchthal and Lindhard).

[1] In column 4, opposite *A* and *B*, the assumption of 2 end-plates per muscle fiber and 8 synaptic endings per postganglionic cell body were made. See references 13–22.

NEUROMUSCULAR BLOCK

The adequacy of the two hypothetical transmitters of the influence of the nerve to the end-plate is illuminated by a consideration of the varieties of neuromuscular block which have been described. For a given nerve impulse arriving at the end-plate, transmission is ordinarily quantal; a muscle impulse is either set up, or not (23, 24). The events which occur during the transmission are for the most part graded. It is to the changes in these graded variables which we must look for the causes of neuromuscular block.

When neuromuscular transmission fails to occur, according to the current concept of transmission, one of three mechanisms is likely to be at

therefore not excited. And third, despite a normal end-plate response, the muscle may fail to respond because its excitability is low or its ability to conduct is impaired. Examples of all three varieties of neuromuscular block have been described.

Curare and other blocking agents. When the application of a chemical substance leaves excitability and conduction in nerve and muscle unimpaired, but prevents the response of muscle to motor nerve impulses, the agent has generally been described as 'curarizing' the neuromuscular synapse, since this is what curare does. It is important to consider, therefore, whether all such 'curarization' represents the same type of neuromuscular block.

The fundamental change which accounts for

the neuromuscular block produced by carare itself is a decrease in the end-plate potential (table 2) (25–27). Transmission fails when the end-plate potential falls below a critical level. The deficiency could be due to a decrease in the amount of transmitter delivered or to a failure of the end-plate to respond to normal transmitter. If the transmitter from the nerve is electric, curare might conceivably decrease the action potential of the nerve. But curare does not affect the action potential of nerves. Or it might conceivably decrease the response of the end-plate mechanism to electric current. In the absence of evidence of electrical excitability of the end-plate, the latter mechanism cannot be tested.

TABLE 2. EFFECTS OF THREE NEUROMUSCULAR BLOCKING AGENTS ON THE FACTORS IMPORTANT IN TRANSMISSION

	CURARE	POTASSIUM	ACETYLCHOLINE ETC.
1. Transmitter			
Action current	0	?	0
Acetylcholine	0	?	?
2. Endplate			
Membrane potential	0	−	−
'End-plate potential'	−	−	−
3. Muscle fiber			
Excitability	0 (−)	−	+
Conduction	0	−	0 (−)

[1] Minus indicates that the agent above causes a decrease or impairment of the variable listed on the left. Zero indicates no effect. Question mark indicates insufficient data.

The evidence that curare does not interfere with the release of acetylcholine caused by nerve impulses has not been invalidated. If acetylcholine is the transmitter, the mode of action of curare is clearly established. Acetylcholine depolarizes the end-plate region, and curare prevents this depolarization by competitive inhibition. This is by far the simplest and most acceptable explanation of the mode of action of curare available. The facts upon which it is based represent strong support for the chemical nature of the transmitter. According to this view, curare acts on the second mechanism of neuromuscular block noted above, that is, it prevents the end-plate from responding normally. With high doses of curare, the electric excitability of the muscle is decreased. But with doses just sufficient to block transmission, the failure of the muscle to respond does not depend on a change in electrical excitability (28, 29).

The mechanisms of various kinds of decurarization contribute to our understanding of curarization. If the curare is eliminated—that is, excreted, or inactivated, or washed out—the neuromuscular transmission is reestablished. The decurarization by azo dyes has been attributed to a special mechanism of inactivation of curare. It is proposed that curare is coupled to the diazo compound and that the coupled curare is no longer able to react with receptive substance (30).

Decurarization may be effected by catelectrotonus applied to the innervated part of a muscle (31). This decurarization might perhaps be partly due to an increase of the end-plate potential produced by the passage of current through this region. More reliance can be placed on the prediction that the curarization will be shown to depend upon a decreased threshold of the muscle fiber adjacent to the end-plate, which permits excitation of the muscle by a small end-plate potential.

The decurarization by cholinesterase inhibitors is associated with a growth in the amplitude and duration of the end-plate potential. The inadequate end-plate potential is rendered adequate, and transmission is reestablished. It is generally agreed that the effect of cholinesterase inhibitors on the end-plate potential is due to the protection from hydrolysis of the acetylcholine produced by the nerve. The phenomena do not exclude a possible excitatory role of the action potential of the nerve upon the end-plate.

When partial or complete curarization to infrequently repeated shocks is established, tetanic stimulation decurarizes. Two separate phenomena are probably present here. One is the summation of end-plate potentials corresponding to succeeding nerve impulses, with the consequent reestablishment of transmission from end-plate to muscle. The other is post-tetanic decurarization (32). Since this outlasts the stimulation of the nerve for a considerable period, its dependence upon the transmitter of the nerve has been denied and less transient agents have been invoked to account for it. It has been argued that the post-tetanic phenomena, including decurarization, result from the potassium liberated during the tetanic activity of the nerve.

Injected potassium can indeed decurarize. In a recent investigation the effects of intra-arterial injections of potassium upon transmission in the curarized frog muscle were observed (33). When potassium was rapidly swept into the muscle in this way, transmission was first facilitated and then inhibited. The facilitation was accompanied by a fall of the threshold of excitation of the muscle and also by a rise in the height of the end-plate potential. An antagonism between curare and potassium has recently been demonstrated on the isolated rat diaphragm (34). Partial neuromuscular block was established by means of curare. The mechanical response to nerve stimulation was used to quantify the degree of block. The necessary concentration of curare was determined for different concentrations of potassium. The concentration of curare rose lineally with the concentration of potassium. The slopes of the curves were different at different temperatures. Extrapolation of these lines to subnormal levels of potassium concentration in the medium showed that at all the temperatures, approximately zero curare would be necessary to produce neuromuscular block in the absence of potassium. It would be most instructive to know the behavior of the membrane potential, the end-plate potential and the electrical excitability of muscle which accompany these changes. Potassium affects all of these variables.

Injected potassium releases acetylcholine in various tissues. It is not unlikely that potassium affects the transmitter, whether it be acetylcholine or action potential (table 2). The block of transmission by potassium is, like that due to curare, associated with a fall of end-plate potential. The mechanism of this change is not yet clear. But unlike curare, potassium in blocking doses decreases muscular excitability and interferes with conduction. During recovery from potassium block, the end-plate potential was observed to recover at a time when the muscle was still unable to conduct impulses (5). Thus potassium blocks by acting at least on the second and third parts of the synapse, the end-plate and the muscle.

Like potassium, acetylcholine can both stimulate and block. The stimulating depolarization of acetylcholine is, however, prevented by curare, whereas that of potassium is not. No data indicate whether or not applied acetylcholine decreases the output of acetylcholine from the nerve (table 2). Acetylcholine does not affect the action potentials of nerves. Hence, if the transmitter is electric, the reduction of end-plate potential by acetylcholine must be due to an effect on the excitability or responsiveness of the end-plate mechanism. A striking difference exists between the blocking produced by curare and that produced by an excess of acetylcholine: in the case of curare, depolarization does not occur; in the case of excessive acetylcholine, excessive depolarization occurs. When acetylcholine is applied to a single end-plate, depolarization begins. After a brief interval, several muscle impulses are set up. Then, despite continued or increasing depolarization, the muscle no longer responds (5). This may correspond to the accommodation which occurs when catelectrotonic current is applied to the muscle. During this period of depolarization due to acetylcholine, end-plate potentials are reduced (35). With increasing concentrations of acetylcholine the end-plate potential becomes gradually smaller and block occurs. It is not clear to what extent the decrease in end-plate potential, produced by acetylcholine in these experiments, is a function of the membrane potential of the end-plate which is depressed by acetylcholine. If the transmitter is acetylcholine, prior reaction of applied acetylcholine with the receptive substance would be expected to reduce the degree of depolarization contributed by a normal quantum of acetylcholine delivered from the nerve.

Neuromuscular block with acetylcholine occurs when the end-plate potential is slightly greater than that which occurs at block by curare. It may be argued from this that the muscular excitability is slightly increased by acetylcholine. A continuing end-plate depolarization may be responsible for this change in the excitability of the nearby muscle relative to the end-plate potentials.

The failure of transmission, which occurs at low frequencies of stimulation of the nerve in the presence of cholinesterase inhibitors, is probably due to an accumulation of acetylcholine and a consequent continuation of end-plate depolarization. In frog muscle, local contracture in the end-plate region of the muscles has been observed under these circumstances (36). When contracture occurs, conduction along the muscle fiber in this region is blocked. Hence contracture in the end-plate region represents one aspect of neuromuscular block, an example of a deficiency of the third part of the system, the muscle.

One wonders how nicotine, tetramethylammonium, and all their congeners act on these variables of the neuromuscular synapse in producing their

block. Do they, like curare, block by a selective inhibitory action, preventing the mediator from depolarizing the end-plate? Or is their block associated with a more complex effect, like that of acetylcholine? The latter seems more likely. Their action is usually designated as curariform. But unless it is established that their action is identical with that of curare, this term is a misnomer. We have reached the point at which by proper analysis it can be determined how these substances produce neuromuscular block. Until this analysis is made, they should not be called cararizing agents, but, more generally, agents which produce neuromuscular block.

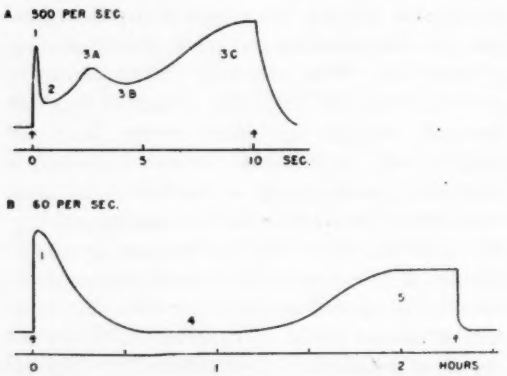


Fig. 2. DIAGRAMS OF MECHANOGRAMS obtained from leg muscles of cats upon stimulation of their motor nerves at rapid frequencies.

The fundamental difference between curare and these other blocking agents seems to be their effect on the membrane potential. Unlike the other agents, curare does not depolarize, at least in normal frog muscle, and it does not ordinarily stimulate muscle. In lizard muscle, however, a depolarization of the end-plate has been reported (4) and, in denervated mammalian muscle, d-tubocurarine is capable of stimulating (37). Denervation renders the muscle much more sensitive to acetylcholine and similar stimulating substances. Evidence from frog single fibers suggests that even in denervated muscles, the end-plate region remains specially sensitive to depolarization and stimulation by acetylcholine (5). One wonders whether, in lizard muscle, curare is capable of stimulating and whether, in denervated mammalian muscle, curare depolarizes.

Other types of block. Other forms of neuromuscular block are seen when the motor nerve is stimulated rapidly for appropriate periods of time. When 300 to 500 maximal stimuli per second are applied to the motor nerve of a leg muscle in the cat, the muscle contracts vigorously but briefly. The characteristic mechanogram is indicated diagrammatically in figure 2A. The successive rises and falls of tension during continued stimulation have been designated the stages of neuromuscular transmission, by numbers, the first rise *stage 1*, the first fall *stage 2*, and the second rise-fall-rise sequence *stages 3a*, *3b*, and *3c*. In the series so far named, there are two which represent failure of neuromuscular transmission, namely, *stage 2* and *stage 3b* (38). During these stages of failure the nerve is able to carry impulses unimpaired. The statistical spike height of a region of the muscle changes proportionately with the tension. Therefore the inference is made that these stages of failure depend on deficiencies at the synapse. It is possible to account for at least one of these stages of neuromuscular failure, *stage 3*. Little more than the simplest description is known about *stage 2*.

In the presence of curare *stage 3b* is diminished. In the presence of cholinesterase inhibitors, it is augmented and occurs at notably lower rates of stimulation. If acetylcholine is injected, *stage 3b* is augmented. The acetylcholine output of excised nerve is highest at first and diminishes with continued stimulation. In the superior cervical ganglion a similar failure occurs; it is affected similarly by acetylcholine, cholinesterase inhibitors, and curare (39). When the ganglion is perfused with eserinized Locke's solution, this stage of failure occurs when the acetylcholine output of the nerve is high and disappears as the acetylcholine output falls. When insufficient substrate for acetylcholine synthesis is available, this stage does not appear in the ganglion. There can be little doubt that *stage 3b* is due to excessive acetylcholine produced by the rapidly repeated nerve impulses. The mechanism of this type of neuromuscular block has already been discussed above. A close correlation of end-plate potential with transmission in these stages has not been made, but it is known that successive end-plate potentials build up in the first seconds of rapid stimulation of the motor nerve, making an increasing degree of end-plate depolarization (40).

When the motor nerve is stimulated at a slower frequency, say 60 to 120 per second, a slower series of changes is apparent (41). The mechanogram is diagrammatically represented in figure 2B. *Stage 1*, the initial rise of tension, is of course present. At this frequency of stimulation the

other early stages are absent. The tension gradually falls as a new and different stage begins, *stage 4*. The tension falls to almost zero and stays there for many minutes. Between *stage 4* and *stage 3b*, several significant contrasts are found. Transmission in *stage 3b* is improved by curare; but transmission in *stage 4* is made worse by curare. Transmission in *stage 3b* is made worse by acetylcholine and cholinesterase inhibitors, but transmission in *stage 4* is improved by these substances. As contraction becomes progressively weaker in *stage 4*, acetylcholine injected during stimulation makes progressively greater increments of contraction. In contrast to *stage 3b*, the acetylcholine content of the motor nerve is low in *stage 4* (42, 43). A similar stage of failure occurs in autonomic ganglia. As transmission across the ganglionic synapses progressively fails, the output of acetylcholine released into the perfusing solution progressively declines (44). When the perfusing solution lacks substrates for acetylcholine synthesis, *stage 4* occurs sooner; and if substrate is added, the transmission is temporarily restored. It seems certain that the failure of transmission in *stage 4*, both in the ganglion and in the neuromuscular synapse, is due to a decrease in the output of acetylcholine from the presynaptic nerve. On the other hand these frequencies of stimulation produce no appreciable fatigue in the action potential of the nerve. These facts represent strong support of the essential role of acetylcholine as the transmitter.

If stimulation is continued during *stage 4*, the tension finally rises again and may eventually reach 60 per cent of the height of *stage 1*. This reestablishment of transmission has been called *stage 5* (fig. 2). The tension may remain at this level for hours. *Stage 5* has been attributed to a gradual increase in the acetylcholine output of the motor nerve. When a nerve is rapidly stimulated for 30 minutes, it loses much of the acetylcholine it originally possessed, but in the succeeding minutes of rest, it synthesizes acetylcholine to a higher level than it originally had. The longer the nerve is stimulated, the greater is this increased rate of restorative synthesis. If nerves taken during *stage 4* and *5* are frozen without the opportunity of restoring their depleted acetylcholine, it is found that nerves contain more acetylcholine in *stage 5*, when transmission is being reestablished, than in *stage 4* when transmission is absent. The action potential of the nerve and the excitability of the muscle do not change as *stage 5* supersedes *stage 4*. The behavior of the end-plate potential has not been examined. Thus the failure of transmission in *stage 4* depends primarily on a deficiency of the output of acetylcholine from the nerve and the restoration of transmission in *stage 5* depends on an increase in the output of acetylcholine. Low calcium and procaine are other agencies which also appear to diminish the acetylcholine output of the nerve; the neuromuscular block produced by these substances is associated with changes in the other phases of transmission as well (45).

Failure of neuromuscular transmission has been attributed to failure of production of acetylcholine by the nerve also in Wallerian degeneration (46–48). The phenomena are parallel in the neuromuscular synapse and the autonomic ganglia. When the presynaptic nerves are cut, little change occurs in the function of the axons peripheral to the cut or in transmission, within the first day. On the second and third days there occurs a progressive failure of transmission. By the end of the third day transmission is absent. Yet the action potential of the nerve remains unaffected at this time and for more than a day after transmission has failed completely. The acetylcholine content, release, and synthesis of these nerves, however, is parallel to the transmission; it is normal during the first day and falls rapidly during the next two days. The cholinesterase content of the synaptic region and the thiamine output of the nerves decline at a much slower rate. From the present evidence, this kind of neuromuscular failure seems to be due in large part to changes in the first phase of the synaptic series, the production of transmitter by the nerve. What happens to end-plate potential and muscular excitability is unknown.

ADEQUACY OF THE HYPOTHETICAL TRANSMITTERS

The adequacy of the two rival transmitters of the influence of the nerve on the end-plate, the action potential and acetylcholine, is summarized in table 3. A plus indicates that behavior of the respective transmitter is adequate to account for the behavior of the synaptic mechanism. A minus indicates inadequacy. A question mark indicates insufficient data on which to decide. The plus in parentheses opposite number 3 under acetylcholine refers to the facts that we know from the effect of cholinesterase inhibitors on the end-plate potential that acetylcholine is present at the peak of the end-plate potential, but we don't know that it is present earlier. The parentheses around the plus under acetylcholine

opposite number 6 refers to the fact that direct data on reduction of acetylcholine output by low calcium exist only for autonomic ganglia.

Acetylcholine is sufficient to account for most of the phenomena. For some of them, for example, *stage 4*, and the transmission failure of Wallerian degeneration, one might almost say that acetylcholine is necessary to account for the phenomena. For the other tests, some other transmitter might suffice. Evidence for such transmitters at this synapse simply does not exist, but new evidence for some other transmitters could make acetylcholine unnecessary to account for the phenomena.

TABLE 3. ADEQUACY OF THE TWO HYPOTHETICAL TRANSMITTERS OF THE INFLUENCE OF THE NERVE TO THE END-PLATE[1]

TESTS	ACTION CURRENT	ACETYLCHOLINE
1. Depolarizes endplate	?	+
2. Quantity produced	?	−
3. Rate of appearance	+	(+)
4. Rate of disappearance	−	+
5. Effect blocked by curare	?	+
6. Quantity decreased by calcium	?	(+)
7. Stage 2	?	−
8. Stage 3b	?	+
9. Stage 4	?	+
10. Stage 5	?	+
11. Wallerian degeneration	?	+
12. Post-tetanic phenomena	−	−
+	1	7+ (2)
?	9	0
−	2	3

[1] + = adequate; ? = unknown; − = inadequate

An experiment reported by deCastro will illuminate this possibility (49). It involves, not skeletal muscle, but the somewhat similar superior cervical ganglion. De Castro cut the cervical sympathetic nerve and induced regrowth to the ganglion of fibers from other cut nerves. He was able to find functional innervation of the ganglion cells from a number of nerve trunks. One such nerve trunk was the vagus. If the vagus is cut central to the nodose ganglion, the efferent fibers in the peripheral trunk degenerate. The afferent fibers, whose cells are in the nodose ganglion, do not degenerate, but start growing at their cut ends. De Castro was able to establish functional connections between these regrowing afferent nerve fibers from the nodosum and the cells of the superior cervical ganglion. Now the nodosum is similar to dorsal root ganglia, and dorsal root fibers are not cholinergic: they do not contain acetylcholine (50). If those afferent fibers of the vagus which establish functional connections with the superior cervical ganglion could be clearly established as not containing or releasing acetylcholine, the transmission in this re-innervated ganglion would occur in the absence of acetylcholine as transmitter. Acetylcholine would not be necessary for the transmission. Some other transmitter would be sufficient. The data upon which such a conclusion could be made have not yet been demonstrated in the re-innervated ganglion. Certainly they have not been demonstrated in skeletal muscle. Hence we can still rely on the conclusion that acetylcholine is the transmitter. The possibilities brought up by the experiment of De Castro represent a challenge to a renewal of the controversy.

REFERENCES

1. COUTEAUX, R. *Revue Canadienne de biol.* 6: 563, 1947.
2. GRAHAM, J. AND R. W. GERARD. *J. Cell. Compar. Physiol.* 28: 99, 1946.
3. LANGLEY, J. N. *J. Physiol.* 33: 374, 1905.
4. BUCHTAL, F. AND J. LINDHARD. *Acta Physiol. scand.* 4: 136, 1942.
5. KUFFLER S. W. *J. Neurophysiol.* 6: 99, 1946.
6. VAN MAANEN, E. F. Unpublished observations.
7. STANBURY, J. B. *Federation Proc.* 7: 257, 1948.
8. HARVEY, A. M. Personal communication.
9. KUFFLER, S. W. *J. Neurophysiol.* 7: 17, 1943.
10. VALENZUELA, F., F. HUIDOBRO AND R. VALDES. *Proc. Soc. Exp. Biol. Med.* 57: 31, 1944.
11. NACHMANSOHN, D. *Annals N. Y. Acad. Sci.* 47 (4): 395, 1946.
12. FESSARD, A. *Annals N. Y. Acad. Sci.* 47 (4): 501, 1946.
13. DALE, H. H., W. FELDBERG AND M. VOGT. *J. Physiol.* 86: 353, 1936.
14. FELDBERG, W. AND A. VARTIAINEN. *J. Physiol.* 83: 103, 1934.
15. LISSAK, K. *Am. J. Physiol.* 127: 263, 1939.

16. SCHEINFINKEL, N. *Helv. Physiol. Acta.* 1: 149, 1943.
17. MURALT, A. v. *Arch. ges. Physiol.* 245: 607, 1942.
18. LANGWORTHY, O. *J. Comp. Neurol.* 36: 273, 1924.
19. BILLINGSLEY, P. R. AND S. W. RANSON. *J. Comp. Neurol.* 29: 359, 1918.
20. GASSER, H. S. AND J. ERLANGER. *Am. J. Physiol.* 80: 522, 1927.
21. CLARK, D. A. *Am. J. Physiol.* 96: 296, 1931.
22. KATZ, B. AND S. W. KUFFLER. *J. Neurophysiol.* 4: 209, 1941.
23. ASMUSSEN, E. *Skand. Arch Physiol.* 70: 233, 1934.
24. STEIMAN, S. E. *Am. J. Physiol.* 140: 269, 1943.
25. ECCLES, J. C., B. KATZ AND S. W. KUFFLER. *J. Neurophysiol.* 4: 362, 1941.
26. ECCLES, J. C. AND S. W. KUFFLER. *J. Neurophysiol.* 4: 402, 1941.
27. KUFFLER, S. W. *J. Neurophysiol.* 5: 18, 1942.
28. ROSENBLUETH, A. *Am. J. Physiol.* 129: 22, 1940.
29. KUFFLER, S. W. *J. Neurophysiol.* 8: 78, 1945.
30. KENSLER, C. J. *Federation Proc.* 8: 233, 1948.
31. KATZ, B. *J. Physiol.* 95: 286, 1939.
32. BROWN, G. L. AND U. S. v. EULER. *J. Physiol.* 93: 39, 1938.
33. WALKER, S. M. AND Y. LAPORTE. *J. Neurophysiol.* 10: 79, 1947.
34. QUILLIAM, J. P. AND D. B. TAYLOR. *Nature* (London) 160: 603, 1947.
35. FILLENZ, M. AND M. HANAFIN. *J. Neurophysiol.* 10: 189, 1947.
36. FENG, T. P. *Chinese J. Physiol.* 12: 331, 1937.
37. MCINTYRE, A. R., R. E. KING AND A. L. DUNN. *J. Neurophysiol.* 9: 297, 1945.
38. ROSENBLUETH, A. AND W. B. CANNON. *Am. J. Physiol.* 130: 205, 1940.
39. BROWN, G. L. AND W. FELDBERG. *J. Physiol.* 88: 265, 1936.
40. SCHAEFER, H. AND P. HAASS. *Arch. ges. Physiol.* 242: 364, 1939.
41. CANNON, W. B. AND A. ROSENBLUETH. *Am. J. Physiol.* 130: 219, 1940.
42. ROSENBLUETH, A., K. LISSAK AND A. LANARI. *Am. J. Physiol.* 128: 31, 1939.
43. LISSAK, K. E., K. NAGY AND J. PASZTOR. *Arch. ges. Physiol.* 245: 783, 1942.
44. KAHLSON, G. AND F. C. MACINTOSH. *J. Physiol.* 96: 277, 1939.
45. JACO, N. T. AND D. R. WOOD. *J. Pharmacol. Exper. Therap.* 82: 63, 1944.
46. FELDBERG, W. *J. Physiol.* 101: 432, 1943.
47. MURALT, A. v. AND G. V. SCHULTHESS. *Helv. Physiol. Acta.* 2: 435, 1944.
48. MURALT, A. v. AND F. WYSS. *Helv. Physiol. Acta.* 2: 445, 1944.
49. DE CASTRO. *Trabajos del Instituto Cajal para investigaciones biologicas* 34: 217, 1942.
50. FELDBERG, W. AND T. MANN. *J. Physiol.* 104: 411, 1946.

PHYSIOLOGY OF NEURO-MUSCULAR JUNCTIONS: CLINICAL ASPECTS

A. M. HARVEY

From the Johns Hopkins Hospital, Baltimore, Maryland

The clinician as well as the physiologist has been stimulated to renewed activity in this field. Crucial experiments of a physiological nature rarely can be performed in human subjects, but the clinical investigator, aided by the recent technical advances in the field of neurophysiology, is now in a position to provide more quantitative studies in human subjects which the experimental physiologist must be able to account for within the scope of his interpretation of the mechanisms normally concerned in neuromuscular and central nervous system function. The great variety of disease processes in man provides a host of ready-made experimental situations in which nature by altering some phase of the functional process has created a test object which the ingenuity of the experimental physiologist frequently cannot duplicate. Recent advances in biochemical methods and in neurophysiological technics have opened a broad field for investigation of diseases of the nervous system as well as the various factors in the normal human subjects which must be maintained under perfect control if the level of neuromuscular and central nervous sytem excitability which means normal function is to be maintained. One cannot hope in the time allowed, nor does it seem desirable, to try and review the clinical aspects of this field in a comprehensive manner. It seems more appropriate to draw on certain examples which illustrate the unique rôle which clinical investigation may play in this realm of physiology and to point out briefly certain recent advances in technic which increase the opportunities for study.

In order to analyze the events taking place when the nerve impulse excites the muscle fiber at the motor end-plate, one must have an accurate estimate both of the number of nerve fibers stimulated and of the number of muscle fibers which respond to each nerve stimulus. In animal experimental work this is done by stimulating the motor nerve with a strength of current sufficient to cause all of the nerve fibers to respond to each stimulus, but not strong enough to cause repetitive firing of the nerve fibers to a single stimulus. The muscle action potential recorded by placing one lead in the tendon and the other in the belly of the muscle remains reasonably constant when the temperature and initial tension of the muscle are kept at a fixed level and serves as an index of the number of fibers concerned in the contraction. It has been shown possible to duplicate the conditions of animal experiments by studying the responses evoked in human muscle by slightly supermaximal stimulation of the nerve supplying that muscle (5). Muscle action potentials in a normal subject were recorded from the hypothenar eminence. In successive records the strength of the ulnar nerve stimulus was increased by 10 per cent. One record showed the development of the maximal potential and further increases in stimulus strength produced no further increase in this muscle action potential.

In order to test the validity of this method for obtaining reliable qualitative and quantitative data in the study of neuromuscular function in human subjects, the next step was to try and reproduce in detail certain well documented animal experiments. The action of curare seemed suitable for this purpose. In a study of the effects of small doses of curarine on neuromuscular conduction in the cat, Brown found when two maximal nerve volleys were set up at varying time intervals that, as the volleys became separated by a period of about 50 milliseconds, there developed an increasing diminution in the response to the second volley which reached its maximum at intervals of between 200 and 400 milliseconds. At that interval the second volley might evoke a response only one-half the size of the first. During a tetanus there was a progressive reduction in the size of the first 10 responses which showed an exponential decline to 25 per cent or less of the original potential.

Following the intravenous injection of tubocurarine, effects were observed in the human subject using this method of study which closely paralleled those observed by Brown in the partially curarized animal. The potential size decreased in proportion to the level of curarization and a single stimulus was followed by a long-

lasting depression similar in its degree and time course to that seen in the cat. When the nerve was stimulated repetitively at a rate of 10 to 15 per second, the successive spike heights showed a progressive decline to less than 50 per cent of their original value. This seemed to provide rather convincing evidence of the value of this type of recording for the study of neuromuscular phenomena in the human subject.

It has been recognized for many years that the distribution of the muscle weakness and the clinical characteristics of the disease known as myasthenia gravis resembled the changes produced in the experimental animal by the administration of curare. The additional knowledge that curarizing substances when given in small amounts to patients with myasthenia gravis greatly exaggerated the symptoms and the fact that the neuromuscular disturbance in this disease was alleviated by the administration of neostigmine, a drug which was known to antagonize the effects of curare, suggested that this resemblance of myasthenia gravis to partial curarization was more than superfiical. It seemed desirable, therefore, to see whether more detailed and quantitative data could be obtained in support of this suggested resemblance.

Facilitation and depression at the neuromuscular junction have been described in animal experiments during partial block induced by small doses of curare. Bremer and Holmes working with partially curarized frogs gave a detailed description of what was called *addition latente*. Each nerve impulse arriving at the neuromuscular junction leaves behind it a process susceptible to summation so that with repeated impulses a certain liminal value is obtained which excites the contraction of more muscle fibers than are excited by a single, maximal nerve volley. Brown extended these observations and showed that this period of facilitation lasted for 50 to 60 milliseconds and was followed by a period of depression of neuromuscular conduction which lasted for more than a second. Thus, there are two possibly antagonistic processes set up by activity at the neuromuscular junction, the one tending to facilitate the passage of subsequent impulses and the other of longer duration having the opposite effect. Both of these phenomena can be demonstrated in the myasthenic patient by recording the action potentials of the muscle abductor digiti quinti evoked by supramaximal paired stimuli to the ulnar nerve (6). A patient who had myasthenia gravis had had no neostigmine for a period of 53 hours. In the first record the second response was 19 per cent greater than the first. At the second interval it was 15 per cent greater than the first and in the third record at 70 milliseconds the second response was 99 per cent of the first. At a further interval of 160 milliseconds, the second response was much smaller than that following the first stimulation. In other words, the second of a pair of maximal stimuli evokes a greater response when it follows the first in less than 70 milliseconds, but at longer intervals a depression of the second response appears. The other point which such records demonstrate is that in certain cases of myasthenia gravis there is a partial block in neuromuscular transmission similar to that seen in partial curarization. This partial block allows one to demonstrate these phenomena of facilitation and depression which probably occur normally following activity at the neuromuscular junction.

In very severe cases the degree of neuromuscular block is greater. The phenomenon of depression following the passage of a single impulse increases in degree and masks the facilitation process. The following illustrates the results in a patient with severe myasthenia gravis in whom neostigmine had been withheld for 72 hours. When the stimuli were 64 milliseconds apart there was a very profound depression of the second muscle action potential. Neostigmine administration increased slightly the response to a maximal stimulus and decreased the depression of the second response to a pair of stimuli but it was not capable of bringing about anything like a normal state of neuromuscular function.

Eserine and its analogue neostigmine have several characteristic actions upon the neuromuscular mechanism. In eserinized muscle the tension of the twitch in response to a single maximal motor nerve stimulus is greatly enhanced. The muscle action potential under these circumstances is converted into a short asynchronous salvo indicating a repetitive response of the muscle fibers to the single stimulus. Furthermore, the excitation of eserinized muscle by a train of motor nerve volleys occurring at a frequency greater than 6 per minute results in a progressive depression of the twitch tension. Stimuli applied to the motor nerve at a rate of 50 per second cause a brief contraction followed by relaxation of the muscle. That a depression of neuromuscular conduction results from the accumulation of a paralyzing concentration of acetylcholine is suggested by the observation that the

enhanced responses of eserinized skeletal muscles to low frequency single nerve volleys are abruptly depressed by the concurrent intra-arterial injection of a minute amount of acetylcholine.

It seemed of interest to study the effects of these drugs in normal human subjects and in patients with myasthenia gravis using the brachial artery as the route of administration and studying the function of the muscle abductor digiti quinti in response to stimulation of the ulnar nerve. The repetitive response described in the experimental animal can be reproduced with ease in the human subject. Here one sees the repetitive discharge evoked by a single motor nerve volley after the injection of 1 mg. of neostigmine into the brachial artery. The voltage of the initial potentials is the same before and after the drug was given. Perhaps the most striking effect following intra-arterial injection of neostigmine was demonstrated by giving pairs of supramaximal motor nerve stimuli separated by varying intervals of time. When the time interval between these paired stimuli was sufficiently brief, the muscle action potential evoked by the second stimulus was significantly lower in voltage and shorter in duration. At the earliest interval at which the two responses could be clearly differentiated the depression of the second potential was greatest, and with increasing intervals it rapidly became less, apparently being indirectly proportional to the time interval separating the two volleys. This depression following neuromuscular activity in the presence of neostigmine has an entirely different time course from that seen in the partially curarized subject or in the patient with myasthenia gravis.

Myasthenia gravis is a disease which is characterized by the reversibility of the process either spontaneously or following the administration of certain chemical substances. It may now be of interest to demonstrate briefly what changes take place when a remission is induced in this disease (7). The following illustration is from a patient with severe myasthenia gravis who had a remission following the removal of the thymus gland and shows two of the changes which occurred. The voltage of the muscle action potential in response to a single maximal motor nerve stimulus greatly increased. In addition, the ability to induce repetitive discharge following the injection of neostigmine into the brachial artery appeared. In these instances it could not be brought about before operation with doses up to 3 mg., while following operation it was easily induced with 1 mg. In another patient prior to thymectomy, without neostigmine, there was a marked depression of the transmission curves. Slight improvement was evident after the administration of neostigmine but the slope of the curve was the same. After thymectomy, before neostigmine administration, a depression of the second response was still present but was now of minimal degree. After the administration of neostigmine there was a different slope to the curve approaching that seen in the normal subject.

Di-isopropylfluorophosphate proved to be a very interesting drug for study in view of its peculiar ability to inactivate cholinesterase irreversibly. When injected in small amounts into the brachial artery of normal human subjects the effects on neuromuscular transmission, as measured by the electrical response of the muscle to maximal nerve stimuli, is essentially similar to that seen following a similar administration of neostigmine (4). In the electromyogram of a normal subject before and after the intra-arterial injection of 0.5 mg. of DFP, one sees the development of repetitiveness to a single stimulus, depression of the second response to two stimuli and the varying responses to a train of stimuli. Thirty-five minutes after the injection of DFP, 9 units of curare were injected into the same artery. The repetitive response to single stimuli disappeared completely and had not returned 30 minutes later. The voltage of the response to a single stimulus was reduced immediately following the curare, but during the period of observation slowly returned to the normal voltage. When virtually normal voltage had returned the repetitive response was still absent. This drug is likewise capable of repairing the defect in neuromuscular conduction in myasthenia as shown in the following illustration. In this experiment one was able to demonstrate by the additional injection of neostigmine a deterioration of neuromuscular conduction with the development of repetitive discharge similar to that seen in normal subjects. The train of events represented might possibly be interpreted to result from the additive anticholinesterase effects of an initial dose of DFP followed by a subsequent dose of neostigmine which permitted the development of a paralyzing concentration of acetylcholine at the neuromuscular junction.

In the experiments in which DFP was injected intra-arterially it was noted that the effect was appreciably less in the patients who had received neostigmine shortly before. Also it was observed

that when DFP was given intramuscularly to the myasthenic patient, who was receiving his regular neostigmine medication, the effect of the DFP was reduced. These observations suggested that neostigmine inhibited the effect of DFP. To analyze this relationship in more detail, 2 patients with myasthenia gravis were given on five occasions an injection of DFP into the brachial artery at a time when they had received no neostigmine during the preceding 24 hours. Injections of DFP were repeated after an interval of one week in the opposite brachial artery at a moment when the patients were at the point of maximal benefit from an intramuscular injection of 2 mg. of neostigmine. When the patient had had a recent injection of neostigmine the intra-arterial injection of DFP did not produce the lasting increase in muscular power which followed the administration of DFP alone. These observations suggest that neostigmine and DFP may compete for a common site of action and that when neostigmine is present at this site it blocks the action of DFP.

Many attempts have been made to demonstrate a chemical component involved in synaptic transmission in the central nervous system. No crucial evidence has yet been introduced. Because of the dramatic effects of DFP on the cholinesterase activity in the central nervous system the following experiments done in normal human subjects are of some interest (3). The daily intramuscular injection of DFP in normal subjects, who were not receiving any other medication, usually resulted in the development of symptoms referable to the central nervous system including excessive dreaming, insomnia, restlessness, increased tension, emotional lability, subjective tremulousness, nightmares, headache, increased libido, giddiness, drowsiness, paresthesias, mental confusion, visual hallucinations and pains in the legs of sciatic distribution. These symptoms were not significantly affected by the administration of neostigmine, but were diminished to some degree by the administration of atropine. Electroencephalograms taken during the period in which these symptoms were present showed an increase in the potential size and in the frequency and irregularity of the rhythm. In many instances there was the appearance of abnormal waves similar to those seen in patients with grand mal epilepsy. These electro-encephalographic changes were promptly reversed by atropine and were not affected by neostigmine or curare. All one can say is that in these experiments central nervous system effects have been produced by di-isopropyl-fluorophosphate which to date has been shown to have no other important action than its ability to destroy cholinesterase. The production of these central nervous system effects by a compound which has been shown to inhibit cholinesterase within the brain suggests that the acetylcholine cycle does play a positive though as yet undefined rôle in central neural function. The point which I wish to emphasize here is that in the human subject one is able to study the effects of this drug without the introduction of any possible conflicting factors such as anesthesia or any operative procedure and in the normal unanesthetized human subject one is able to record the subjective alterations in normal central nervous system behavior as well.

At this point I would like to review very briefly certain additional technics by which the clinical neurophysiologist may be able to obtain more detailed knowledge of central nervous system function and to study in a more precise fashion the effects of various diseases and chemical agents on this function.

Piper in his monograph on *Electrophysiology of Human Muscle* was the first to suggest that conduction velocity in a peripheral nerve may be measured by stimulating the nerve at two different sites, measuring the latencies of the muscular responses evoked by each shock and dividing the latency difference by the distance between the points of stimulation. In recent experiments Hodes (9) has used this method of studying conduction velocity of the skeleto-motor nerve fibers supplying paretic muscles. The electrical studies consisted of the percutaneous application of single supra-maximal shocks to the motor nerve and the recording of the muscle action potentials thus produced, as described earlier in this discussion. Conduction velocity was measured by the method described by Piper. Hodes obtained action potentials from some 500 paretic muscles of poliomyelitis patients and expressed them as percentage of the voltage of corresponding normal muscles and compared them with the muscle strength estimated by clinical testing. Such data showed that the size of the action potential indicated the severity of the muscular weakness as had been demonstrated in cases of peripheral neuritis by Dr. Kuffler and myself at an earlier date (8). Hodes then plotted the maximal conduction velocity of the nerve to a paretic muscle against the maximal voltage of that muscle in over 50 cases. It was evident from these results that if a given muscle of one patient were stronger than the correspond-

ing muscle of another patient then the nerves supplying the former would conduct more rapidly than would those innervating the latter. Preliminary histological observations of anterior horn cells of paralyzed monkeys showed results suggesting a neural involvement dependent on axonal size.

Even with the relatively primitive electrophysiological technics available during the early part of this century, clinical physiologists became interested in their application for the study of spinal cord reflex activity in human subjects. Hoffman (10) was among the first to demonstrate the feasibility of studies of this type. He first devised a method for recording the impact of a blow on the patellar tendon and in the same record the electro-myogram from the quadratus muscle. He realized that the measurements were complicated by the characteristics of the sensory end organ. He went a step further and found that by electrical stimulation of a mixed nerve he was able to excite a reflex response which could be recorded electrically. This, as was to be expected, gave a shorter reflex time: 17 to 18 milliseconds for the patellar reflex and 28 to 29 for the achilles reflex.

Dawson and Merton have recently begun studies using this type of experiment for determination of spinal reflex activity. This method should be very fruitful in the study of certain diseases of the spinal cord and also the activity of various pharmacological agents on this region of the nervous system. It is possible to stimulate one motor unit and get a reflex response in the same motor unit.

Dr. George Dawson, working at the National Hospital in London, found in studying a patient suffering from myoclonic seizures that electrical stimulation of a peripheral nerve resulted in changes of electrical potential detectable on the scalp (1). These changes in potential were largest over the hemisphere on the side opposite to that stimulated. They were located near to the midline when the lateral popliteal nerve in the leg was stimulated and more laterally when the ulnar nerve was stimulated in the arm. Extending these studies to healthy subjects, Dawson has been able to demonstrate that potential changes of cerebral origin which probably arise in the central and post-central cortex may also be detected on the scalp following electrical stimulation of peripheral nerves. In a record taken from Dawson's paper which appeared recently one can see the responses to stimulation of the left and right median nerves at the elbow and the left lateral popliteal nerve at the head of the fibula. The electrodes were applied over the surface markings of the sensory motor area. A stimulus to the left median nerve produced a potential change maximum near the electrode on the right side 6 cm. from the midline. A stimulus to the right median produced a similar disturbance on the left side of the head. Stimulation of the left lateral popliteal nerve gave rise to a potential change nearer to the midline electrode than to any of the others. The latency of the response to stimulation in the leg was 36 plus or minus 2 milliseconds which was about 14 milliseconds longer than when the median was stimulated.

Finally, I would like to discuss briefly an interesting group of clinical changes which are associated with alterations in serum potassium concentration. The introduction of the flame photometer has made it possible to study more extensively various conditions associated with changes in potassium. The following illustrates what has been accomplished by improvements in the chemical technics. As one observes the ratio of potassium in serum water and extracellular fluid water calculated by observers over a period of years one can see that it now approaches the theoretical value expected from the Gibbs-Donnan equation (2). It now seems possible that by determinations of serum potassium level and the concentration within the muscle itself one may be able to get accurate values of the relative concentrations bathing the two sides of the cell membrane. These studies when combined with physiological measurements will find useful application in the study of patients with family periodic paralysis having low serum potassium as well as certain patients with diabetic acidosis who also exhibit hypokalemia. A clinically similar paralysis may accompany elevation of serum potassium in certain cases of uremia. It is of interest that in all three of these clinical conditions there is profound muscular weakness with diminution and loss of reflex activity, but little or no alteration in sensation. The distribution of the weakness, particularly marked in the extremities and trunk, is also the same. Preliminary studies have been made of the electromyogram in patients with family periodic paralysis during an attack and in the normal state. It will be noted than when the electrodes are placed far apart along the muscle that the potential tends to be more monophasic in character than usual. The normal diphasicity is maintained when the electrodes are placed together in the middle third of the muscle. It has been noted that there

is very little change in the shape of the potential where the electrodes are placed 2 centimeters apart. Although there are other possible explanations this type of experiment suggests that the defect may be in the transmission of the excitatory wave along the muscle fiber itself. Similar studies are not available in the clinical states associated with elevation of serum potassium.

Thus, one can see that although the methods are still relatively crude and that the potential contributions of the clinical neurophysiologists are still in the developmental phase it is becoming possible to gather more and more reliable facts and the unique physiological experiments created by naturally occurring disease may uncover many stimulating leads.

REFERENCES

1. Dawson, G. D. *J. Neurol. Neurosurg. Psychiat.* 9: 5, 1947. *Ibid.* 10: 137, 1947.
2. Folk, B., K. L. Zierler and J. L. Lilienthal Jr. Unpublished observations.
3. Grob, D., A. M. Harvey, O. R. Langworthy, and J. L. Lilienthal, Jr. *Bull. Johns Hopkins Hosp.*, 81 (4): 257, 1947.
4. Harvey, A. M., J. L. Lilienthal, Jr., D. Grob, B. F. Jones and S. A. Talbot. *Bull. Johns Hopkins Hosp.* 81 (4): 267, 1947.
5. Harvey, A. M. and R. L. Masland. *Bull. Johns Hopkins Hosp.* 68: 81, 1941.
6. Harvey, A. M., J. L. Lilienthal, Jr. and S. A. Talbot. *Bull. Johns Hopkins Hosp.* 69: 547, 1941.
7. Harvey, A. M., J. L. Lilienthal, Jr. and S. A. Talbot. *Jour. of Clinical Investigation* 21: 579, 1942.
8. Harvey, A. M., S. W. Kuffler and J. B. Tredway. *Bull. Johns Hopkins Hosp.* 83: 103, 1945.
9. Hodes, R. The 17th International Physiological Congress, Oxford. *Abstracts of Communications*. P. 190.
10. Hoffman, P. "Untersuchungen über die Eigenreflexe Menschlicher Muskeln. Berlin: Julius Springer, 1922.

AMERICAN SOCIETY OF BIOLOGICAL CHEMISTS

Symposium on Methods for the Determination of the Purity of Substances of Biochemical Interest[1]

VINCENT DU VIGNEAUD, *Chairman*

VALUE OF CHROMATOGRAPHY IN DETERMINING THE PURITY OF SUBSTANCES OF BIOCHEMICAL INTEREST

HAROLD G. CASSIDY

From the Department of Chemistry, Yale University, New Haven, Connecticut

In order to evaluate chromatography as a method for determining purity we must first define what we mean by purity. This will enable us to set up criteria for judging the methods which test purity. Then, after having set the standards we may evaluate the method of chromatography by seeing how effectively it meets these standards. The standards which we set up are those abstracted, as it were, from all methods of testing purity: they are the principles of the methods of testing purity.

Upon close examination of what we mean by the determination of purity we discover that what we really mean is the detection of impurity. There is no sure way of proving a substance to be pure. We can only show with any confidence that if it contains impurities of thus and such a kind they cannot exceed thus and such an amount. If we use crude methods of finding impurities then the degree of purity can be stated only to a crude approximation, and with progressively acuter examination we become correspondingly more able to detect impurity. It is evident, therefore, that the definition of a degree of purity must be made on the basis of the operations used to detect impurity. We do not define 'purity'. We define a degree of freedom from impurity as disclosed by certain given measurements. This operational definition of purity emphasizes that the result of the measurement depends on the instruments used, and emphasizes, further, that purity is a relative matter. There is consequently no such thing as absolute purity, and any statement which purports to define an absolute purity is operationally meaningless. This has been known, or suspected, for a long time, of course, and an excellent discussion may be found in Timmermans' *Chemical Species* (1), and in Barnes' article in a recent symposium (2). When we use the word 'purity' in the chemical sense we must remember that we mean qualified purity, degree of purity.

What, then, are the principles of the methods for determining purity? They are that in all methods of testing purity two operations are involved. All methods of testing purity involve tests of identity, and the most subtle and reliable tests of purity involve also, and usually simultaneously, attempts at separation (3). Consider, for example, the melting point. The melting point is sometimes used as a test of purity, but only as it is used to test identity. It may indicate identity of the one substance with another which *may* previously have been proven pure. For a more profound test which involves the melting point one has to use the cooling curve (4). For if a substance is pure it will pass through a fractional solidification involved in taking a cooling curve in such a way that, barring supercooling, the solidification temperature at which crystal formation first begins will be the same as that at which the last material solidifies. Thus the cooling curve, since it combines an attempt at separation (that is, fractional crystallization) with a test of identity of the putatively purified fractions (that is, freezing temperature) meets the requirements laid down for the most acute type of test of purity. Moreover, such a test of purity is sufficient to itself in that it does not, as a test of purity, require any standard or reference substance (as does the melting point).

The boiling point measurement and the Swieto-

[1] Atlantic City, N. J., March 1948.

slawsky ebulliometry (5) bear the same relation to each other as do the melting point determination and the cooling curve measurement; also the single solvent partition and the Craig distribution measurement (6) bear an analogous relation. The relationship is that between a single step and a countercurrent operation (3).

However, any test of purity is subject to certain limitations. A eutectic mixture behaves like a pure substance in the cooling curve apparatus; an azeotrope behaves like a pure substance in an ebulliometer; and analogous behaviors will be recognized in other separations. It therefore follows that no single test of purity is conclusive; only is the weight of evidence for purity overwhelming when several tests have been applied and found to be met. For example, the ebullioscopic measurement at several pressures is needed to exclude with certainty the existence of an azeotrope; in the case of solvent distribution, the use of several quite different solvent pairs is needed.

The question may now be put: does the chromatographic method meet the requirements for a method of testing purity (2)? In the chromatographic method a mixture to be separated is passed over an adsorbent (fig. 1). If several substances with different adsorbabilities are present the more strongly held one will be retained near the point of application of the mixture, the less firmly held ones further along the adsorbent. Then, by developing the chromatogram, the components may be separated from each other on the basis of their adsorbabilities. This is the principle of chromatography; the various methods of its application are not of consequence here. When the method is used as a test of purity the substance to be tested is applied to the adsorbent. As the substance meets the adsorbent and becomes adsorbed to the surface any impurity will be expected to be either more or less strongly adsorbed, so that whether on development or frontal analysis there will be evidence of more than one zone. If, however, the substance is pure, the chromatography will not separate it, and a single zone will result. That a single zone is present may be established by any of the methods of recognizing zones: concentration measurements in frontal analysis, displacement development, or elution analysis, or by any other methods which can show that one and only one zone is present in the chromatogram (7). The answer to the question must therefore be that chromatography meets the requirements for a method of testing purity, for not only does it involve a test of identity (since different substances would be adsorbed to different degrees), but it involves an attempt at separation by virtue of the countercurrent application of the adsorption partition.

An examination of actual cases needs to be made to see whether in practice the method redeems its promise. Since not very many tests of chromatography have been made for the express purpose of detecting impurity in substances, but rather chiefly for the purpose of separating mixtures, an attempt will be made to derive information about the effectiveness of the chromatographic method from both classes of experiments. The literature on chromatography is voluminous, and even the papers which are directly useful to this argument are too many to deal with here. For this reason only a few which illustrate the use of the method with protein, carbohydrate and lipoidal substances will be chosen.

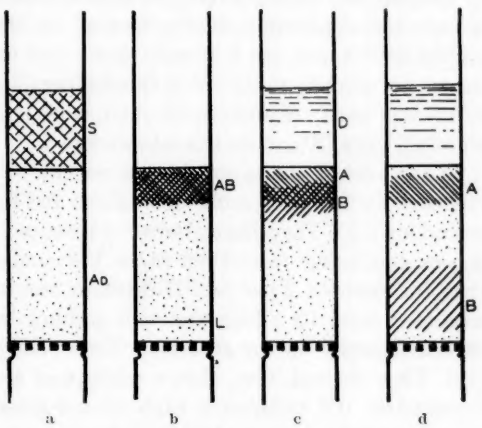


Fig. 1. Chromatography of a mixture of two solutes *A* and *B*. *a*. The solution containing *A* and *B* is placed above the adsorbent AD.; *b*. The solution has been pressed into the adsorbent, a mixed zone *AB* is formed, and the 'front' of pure liquid is shown at *L*; *c*. developer liquid *D* is being passed into the column and development of the zones is beginning; *d*. continued passage of developer liquid has produced a developed chromatogram, with the separate zones shown at *A* and *B*.

The use of chromatography is rather new in the field of the proteins and their degradation products, but already we have excellent evidence of its value. Thus Martin and his coworkers (8), using paper-strip partition chromatography (9), examined an authentic sample of Thudichum's 'glycoleucine' and found it to be leucine. They showed, in another experiment, that not more

than 0.03 per cent of the nitrogen of hydrolyzed spinal cord could be present as *nor*leucine since when 0.5 μg. of *nor*leucine was added to 1.2 milligrams of hydrolysate it could be clearly recognized, but none could be detected in the hydrolysate alone. The work of Moore and Stein (10) at the Rockefeller Institute shows how very accurately the chromatographic method can analyse a protein hydrolysate, and the further sharpening of the chromatographic tool by the use of the radioactive 'pipsyl chloride' developed by Cannan and his coworkers (11) indicates that though the application of the method in the protein field is new, yet it already shows that we can expect much from the use of chromatography as a test of purity of amino acids, and, indeed, of peptides, enzymes and similar substances.

In the field of the carbohydrates we can call immediately upon the results of Wolfrom and his coworkers (12). They have shown, in a series of papers emanating from Ohio State University, that, for example, 2 per cent D-mannitol can be separated from its admixture with sorbitol by chromatography on the adsorbent Florex XXX (13). They showed, also, that a mixture of approximately 0.2 milligrams each of α-d-galacturonic acid, d-galactose, d-glucose, d-xylose, and l-rhamnose could be separated into five zones on the same adsorbent, using isopropyl alcohol as the solvent and developer.

Very many lipoid-soluble substances have been examined chromatographically for purity. Elegant work of this kind was done by Winterstein and Stein (14) who were able to detect and separate 2 milligrams of dipalmityl ketone in 300 of a hentriacontane which showed the theoretical analysis and a good melting point. They also could concentrate 0.5 per cent ergosterol in admixture with cholesterol through chromatography on alumina. It was in this laboratory also (15) that chromatography was used to isolate 0 .05 per cent carbazole from anthracene. The anthracene so obtained showed the blue fluorescence which is known to be quenched by 1/30,000 per cent of naphthacene. The original sample of anthracene had contained 0.5 per cent of naphthacene. Mair (16) and his coworkers at the Bureau of Standards have shown that the aromatic hydrocarbon content of a mixture of paraffins, naphthenes and aromatics can be determined to within ±0.2 per cent. Zechmeister and McNeely (17) showed that it was possible to determine 1 to 2 per cent of *cis*- or *trans*-stilbene in admixture with the other form.

In the fat series Kaufmann (18) showed that the purest grades of caproic, lauric, myristic, palmitic, oleic and stearic acids contained significant amounts of impurities consisting of higher and lower fatty acids. The method was found applicable also to glycerides. Cassidy (19) was able to demonstrate unequivocally the presence of 0.67 per cent ester in the presence of mixed fatty acids by chromatography on charcoal. Claesson (20) could demonstrate 3.9 per cent of lauric acid in admixture with myristic.

Ruzicka and his colleagues (21) have demonstrated, through their precise fractionations of natural mixtures, the value of chromatography in separating and testing the purity of lipoid-soluble substances. As an example of the precision of their work one may cite the isolation of friedelin. From 82.2 gm. of nonsaponifiable material (representing 1500 kg. of pig spleen) they isolated 25.4 mg. of the triterpene friedelin. This represented 0.03 per cent of the nonsaponifiable fraction, and was present as an impurity from cork stoppers into contact with which the material had come.

No discussion of the use of chromatography as a test of purity would be representative without mention of the work on plant and animal pigments. We might mention here the work of Zechmeister (7) and of Strain (7).

Chromatography shows a limitation analogous to that already described in connection with other testing methods, that is, two different substances may be adsorbed under certain conditions to the same extent, thus giving a false impression of purity. This situation is analogous to that encountered in eutectic formation and azeotropism (3). The remedy is the same as in those cases: namely, to change the conditions of the test. With adsorption this involves using different adsorbents and different solvents, and preferably choosing these to have different polarities from those first taken. An observation of Bauernfeind and coworkers (22) may supply an example of this limitation. These workers found two pigments which, as they said, were chromatographically identical by mixed adsorption on activated alumina and development with reagent chloroform, but which were not identical spectrophotometrically.

A further limitation on chromatography, and one not absent from other methods of testing purity, is in effect the reverse of this. Instead of a mixture appearing pure, a pure substance yields several zones. This is often due to some catalytic effect of the adsorbent in which a sub-

stance is altered (23). Thus a pure *cis* compound can be altered to a mixture of *cis* and *trans*, which then yields two zones. This effect can often be anticipated and corrected; fortunately it is not very common. The classical example of this effect was the observation by Gillam and El Ridi (24) that carotenes could be isomerized by chromatographic adsorption. This observation led, in the hands of Zechmeister and his students, to the remarkable and important work on the configuration of the polyene pigments (25).

That biologically important substances may be altered on the chromatographic column has been recognized, and, as an example, has led to the development of analytical methods which would circumvent this difficulty. For example, Haagen-Smit and co-workers (26) developed a shortened method for analyzing provitamin A substances from plants in order to minimize the losses which accompanied certain methods in common use.

Many other phenomena may occur in the column on active adsorbents. As one example we may cite the observation of Trappe (27) that triglycerides may be saponified on Brockmann alumina even with nonaqueous solvents. Trappe also recorded that free as well as esterfied cholesterol underwent catalytic decomposition on silica columns, with the development of a blue-red coloration, when adsorbed from solvents which contained hydrochloric acid (as would chlorinated solvents which had been standing around for some time).

Fortunately, as has been remarked above, these situations are relatively rare, and can be anticipated when very active acidic or alkaline adsorbents are being used. The appearance then of two zones would not necessarily imply impurity in the initial substance.

It must be evident from the remarks made so far that chromatography should be a very useful method for testing purity of substances provided that they are adsorbable, and all substances show this property to some extent. Chromatography may be used to test purity of substances in the gaseous form (20), or as liquids, or in solution, and it applies to adsorption at non-mobile interfaces as well as mobile, that is, to adsorption on solids as well as to adsorption on foams and emulsions (28). Dr. Craig has suggested that I mention that the chromatographic method, perhaps due to the mechanics of the arrangement, is the most *efficient* separation method we have, for a short column may provide separations which indicate it to have many hundreds of 'theoretical plates' (10, 29). When the method is used within its proper sphere it can supply a most powerful tool for testing purity. This is because it meets

TABLE 1. ORDERS OF ABSTRACTION

LEVEL OF ABSTRACTION	LIVING	NON-LIVING	ORDER OF INCREASING ABSTRACTION
	First-Order Facts[2]		
Macroscopic[1]	Organisms	Substances (bulk) (bulk compounds)	↑
Microscopic	Cells	(bulk elements)	
	Inferential Facts		
Submicroscopic	Colloids		
Molecular	Molecules and atoms		
Subatomic	Protons and electrons		

[1] By abstraction is meant a process of leaving out details (30). In passing from one level to another level of increased abstraction there are left out many details which are used to describe the first level. Thus in passing from the subatomic level to the molecular there are left out all sorts of information about nuclear construction, electron shells, electron spins, etc. These may, however, be implied in the symbol used for the molecule or atom, but they cannot be explicitly stated in the terms used for dealing with the more abstract level. This is in part the dilemma of form and function. It is the functioning of protons and electrons which makes the form of molecules. It is the functioning of molecules which makes the form of colloids and cells. There is naturally some overlapping of these levels.

[2] First-order facts are facts which are directly observable. Inferential facts are facts which are inferred from, and to explain, first-order facts.

the requirements which we found to be characteristic of a good method, namely, because it involves an attempt at a separation of the substance under test and at the same time a comparison of the fractions obtained to test their identity.

The problem of the nature of chemical purity has aroused so much discussion that it may not be amiss to attempt here to clarify certain points involved in it. In table 1 is presented a classification of certain phenomena. The concept of chemical purity can be applied correctly *only* at the level of macroscopic and microscopic first-order facts. Pure is a word with many meanings, each depending on the context within which the word is used. It is essential to distinguish the different meanings. Thus when one speaks of a pure cell-suspension the word is being used in an entirely different sense from that used when one speaks of a pure substance. In the former case one can prove that all the cells are of one kind; in the latter it is impossible to prove that all the molecules are identical. The word pure cannot be used at the molecular or subatomic levels with any chemical meaning. There is no such thing as a chemically pure molecule. (And of course, in the case of bulk elements, these are by definition pure. The word 'bulk' is used here to qualify 'element' which is often used to mean 'a molecule of an element'.) Biochemists, since they are working in a borderline field, are beset with all sorts of difficulties in nomenclature. This note is written with the hope that at least one of the difficulties will be clarified. The application of the idea of levels of abstraction can clarify many other problems. An example of the 'paradoxes' which can be devised by dealing in the same terms with two levels of abstraction at once can be found in a recent paper on purity (31).

REFERENCES

1. J. Timmermans. *Chemical Species*, Translated by R. E. Oesper. New York: Chemical Publishing Co., Inc., 1940.
2. R. B. Barnes. *Analytical Chemistry* 20: 96, 1948.
3. H. G. Cassidy, *J. Chem. Education* 23: 427, 1946.
4. E. L. Skau and H. Wakeham, in *Physical Methods of Organic Chemistry*. New York: Interscience Publishers Inc. I: 1, 1946.
5. W. Swietoslawsky, *Physical Methods of Organic Chemistry*. New York: Interscience, I: 47, 1946.
6. L. C. Craig. See contribution to this Symposium.
7. L. Zechmeister and L. Cholnoky. *Principles and Practice of Chromatography*. New York; John Wiley and Sons, Inc., 1942; H. H. Strain. *Chromatographic Adsorption Analysis*. New York: Interscience Publishers, Inc., 1942; T. I. Williams. *An Introddction to Chromatography*. London: *Blackie and Son, Ltd.*, 1946.
8. R. Consden, A. H. Gordon, A. J. P. Martin, O. Rosenheim and R. L. M. Synge *Biochem. J.* 39: 251, 1945.
9. A. H. Gordon, A. J. P. Martin and R. L. M. Synge. *Biochem. J.* 37: xiii, 1943; R. Consden, A. H. Gordon and A. J. P. Martin. *Biochem. J.* 38: 224, 1944. See also A. J. P. Martin. *Ann. New York Acad. Sci.* 49: 249, 1948.
10. S. Moore and W. H. Stein. *Ann. New York Acad. Sci.* 49: 265, 1948.
11. A. S. Keston, S. Undenfriend and R. K. Cannan. *J. Am. Chem. Soc.* 68: 1390, 1946.
12. B. W. Lew, M. L. Wolfrom and R. M. Goepp, Jr. *J. Am. Chem. Soc.* 68: 1449, 1946.
13. V. R. Dietz. *Ann. New York Acad. Sci.* 49: 315, 1948.
14. A. Winterstein and G. Stein. *Z. Physiol. Chem.* 220: 247, 1933.
15. A. Winterstein, K. Schön and H. Vetter. *Z. Physiol. Chem.* 230: 158, 1934.
16. B. J. Mair. *Ann. New York Acad. Sci.* 49: 218, 1948.
17. L. Zechmeister and W. H. McNeely. *J. Am. Chem. Soc.* 64: 1919, 1942.
18. H. P. Kaufmann. *Angew. Chem.* 53: 98, 1940.
19. H. G. Cassidy. Unpublished.
20. S. Claesson. *Arkiv. Kem. Mineral Geol.* 23: #1, 1946; *Ann. New York Acad. Sci.* 49: 183, 1948.
21. V. Prelog, L. Ruzicka and P. Stein. *Helv. Chim. Acta* 26: 2222, 1943.
22. J. C. Bauernfeind, W. Baumgarten and C. S. Boruff. *Science* 100: 316, 1944.
23. L. Zechmeister. *Ann. New York Acad. Sci.* 49: 145, 1948; *reference 7*.
24. A. E. Gillam and M. S. El Ridi. *Biochem. J.* 30: 1735, 1936.
25. L. Zechmeister. *Chem. Rev.* 34: 267, 1944; *Ann. New York Acad. Sci.* 49: 220, 1948.
26. A. J. Haagen-Smit, C. E. P. Jeffreys and J. G. Kirchner. *Ind. Eng. Chem.*, Anal. Ed. 15: 179, 1943.
27. W. Trappe. *Biochem. Z.* 306: 316, 1940.
28. L. Shedlovsky. *Ann. New York Acad. Sci.* 49: 279, 1948.
29. A. J. P. Martin and R. L. M. Synge. *Biochem. J.* 35: 1358, 1941.
30. W. Johnson. *People in Quandaries*. New York: Harper Bros., 1946.
31. H. Eyring. *Analytical Chem.* 20: 98, 1948.

COUNTER-CURRENT DISTRIBUTION

LYMAN C. CRAIG

From the Rockefeller Institute for Medical Research, New York City

The classical approach of organic chemistry to the problem of purity has always been that of fractionation. A preparation has been regarded as 'pure' if all attempts to fractionate it have failed to change its measurable properties. It should be satisfying to the organic chemists to know that this approach has recently been stated (1) by a physical chemist, Professor Eyring, to be the definition of a component in the thermodynamic sense of the phase rule. Irrespective of viewpoints, nearly all fractionation techniques are useful in purity studies.

Biochemistry deals with an enormous number of chemical compounds with widely differing properties. Some are gases, some are liquids and some are solids either crystalline or amorphous. Wide variation in stability is encountered. Obviously no single fractionation technique will be the most effective for the separation of each type. Since the determination of purity requires maximum separating power under conditions of complete stability, a choice of the most effective method must be made. Fortunately, the different methods often supplement each other and wherever possible, more than one method must be applied. Measurement of physical constants with adequate precision is taken for granted, as far as the present discussion is concerned.

In order to choose the most effective fractionating technique we must know something of the nature of the compound, in biochemistry particularly something of the stability of the solutes involved. Unquestionably many compounds of importance have such poor stability except in solution in a certain environment that they can never be crystallized or obtained free of solvent. Yet such a substance could cause a very specific biochemical response and it would therefore be of the greatest interest to know if such a response were caused by a single chemical individual or whether a collection of individuals were responsible for the effect. In such a problem counter-current distribution offers possibilities.

A few years ago a problem of this general character was encountered by us in studies dealing with synthetic antimalarial drugs. Unexpected toxicity was encountered in connection with the clinical study of a particular drug. This raised the question of purity and the chemists were asked for reliable data to prove that they were not at fault in claiming a pure preparation. At the time no satisfactory method existed for demonstrating purity since the properties of the compound did not permit the fractionation techniques available to be applied under favorable circumstances.

An approach to the problem was made through the application of extraction data. Although extraction is one of the oldest procedures of organic chemistry, it is usually applied in such a manner that results only approximate in nature are obtained. Purity studies require quantitative data. Particularly in biological circles it has been appreciated for some time that if enough cases are at hand, the results can be treated statistically and thereby greater precision achieved. Therefore, multiple extractions were performed in counter-current manner so that the steps of the binomial expansion were exactly followed (2). This permitted the mathematics of probability to be applied in the interpretation of the results (3, 4). The name 'Counter-current Distribution' was given to the process and an apparatus was devised whereby hundreds of nearly perfect individual extractions could be performed in an hour or more. The final result of the process could be most conveniently expressed graphically by plotting the fraction of solute in a given extraction cell against the serial number of the cell. This gave a pattern which with a single solute will exactly correspond to the normal curve of error as in figure 1. Figure 1 is the pattern of an actual distribution made with a sample of benzyl-penicillin (5). Since the curve is an averaged or statistical result, a considerable error in a single extraction is unimportant.

The requirements in order to accomplish a normal curve of error are the following: Two phases must be found in which the solute will distribute itself in approximately equal amounts upon equilibration. The partition ratio must be satisfied at every step. The phases must separate easily. The partition isotherm must be linear, i.e.,

the partition ratio not be different with different concentrations of solute. Often the temperature must be held constant. The process is restricted to those cases where the partition ratio really is a constant. Each of the necessary factors can be checked separately or a distribution can be made and from the nature of the pattern it can be derived whether or not they have been met. If the experimental distribution agrees with the normal curve of error, it follows that the necessary conditions have been met and that the solute is a single substance or more than one substance with nearly identical partition ratios. If a deviation occurs as in figure 2, determination of the partition ratios of the solute in the individual tubes where deviation is shown will decide if the cause is impurity.

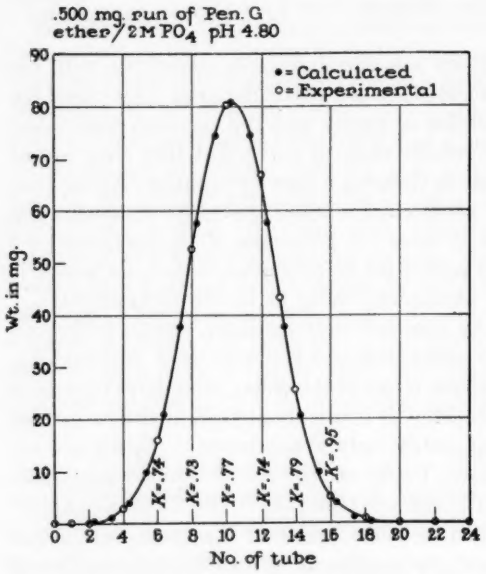


Fig. 1. TWENTY-FOUR TRANSFER DISTRIBUTION of benzylpenicillin.

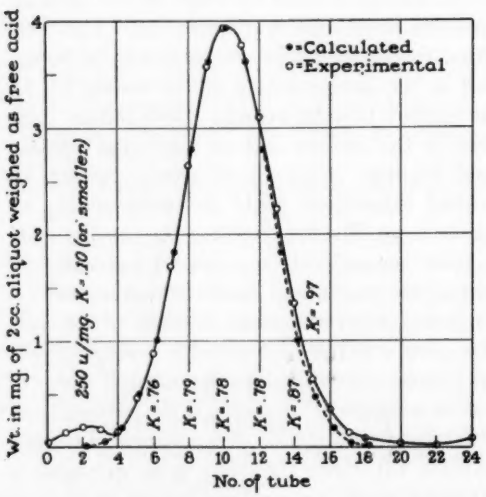


Fig. 2. TWENTY-FOUR TRANSFER DISTRIBUTION of slightly impure benzylpenicillin.

A gradual drift in the ratios as shown in figure 2 definitely indicates impurity. Figure 2 is a distribution with a sample of benzyl-penicillin which contains a small percentage of Δ^2-pentenylpenicillin (5) and a little of another impurity. Experimental achievement of the curve of error indicates purity as far as the method and the particular system is concerned, though it alone is by no means conclusive evidence for purity. The result is analogous to a precise fractional distillation followed by an ebulliometric or boiling point study of each of the fractions. Or it is analogous to fractional crystallization followed by a phase rule study such as the determination of the solubility curve in the solvents used for the recrystallization. However, none of these methods absolutely exclude impurity but for practical purposes, if more than one of the procedures indicates purity or a curve of error is obtained in more than one entirely different system, then the preparation will be of high purity in by far the largest majority of cases.

With any method of fractionation, the degree to which a possible impurity can be excluded or the precision with which it can be revealed will depend upon the intrinsic separating power of the method. Separating power basically depends upon the selectivity of the method and the number of perfect transfers applied. A continuous fractionation process does not have discrete steps and must be interpreted in terms of the 'equivalent of the number of perfect stages.' Often this number is not constant. Stepwise extraction offers a process which can be made to operate at complete equilibrium and with any desired number of extractions. Assumptions are not required. The phenomenon corresponding to 'constant boiling mixtures' is eliminated by working at sufficient dilution. The constant which determines the selectivity is the partition ratio. It is of importance, therefore, to understand the full significance of the ratio.

In more than one instance the partition ratio has been spoken of as depending on 'solubility'. This would appear to be quite erroneous as far as the type of partition ratio used in this connection is concerned. Only at the point of complete saturation of the solute in each phase will the partition ratio correspond to a ratio of solubilities. At concentrations below this, constant pro-

portionality of the activity of the solute in the two phases will seldom be found. It would seem better to consider a partition ratio as a measure of the relative attracting tendency of two phases for a solute at a specified concentration. The basis of the ratio is therefore a subtle one and usually quite different from solubility.

In the technical literature of fractionation (6) the selectivity of a particular extraction system is defined in terms of the ratio of the partition coefficients of two arbitrary solutes. This ratio is called β and equals K_1/K_2 in the particular system. A system in which β is large is a selective system for separating the mixture. Obviously this should be the system of choice for demonstrating impurity if the arbitrary solutes are closely related and are similar to the one on which the purity study is to be made.

In general, those systems will be the most selective in which there is the greatest contrast in the hydrophobic-hydrophyllic solvent properties of the two phases concerned. Where the solute is an acid or base, buffers are especially helpful. As regards more subtle degrees of selectivity, no general rules can be offered. The choice must be based on experimental trial.

Once the selectivity of a system has been established it is a simple matter to calculate the number of transfers required in order to detect a given impurity or to exclude it to a given percentage. Hypothetical distributions such as figure 3 bring out this point. Consider a mixture of 90% *A*, $K = 1$ and 10% *B*, $K = 1.2$, i.e., $\beta = 1.2$. The sum of *A* and *B curves*, the top curve, would differ sufficiently from the curve of error so that a small percentage of *B* in *A* could be detected with 100 transfers. This requires a few hours' time with our present equipment, which includes a distribution apparatus with 54 cells. Two or even three times this number of transfers are within reach. Time does not permit further discussion of this type of calculation. The picture given here is that based on weight alone and weight is certainly the most inclusive analytical method as far as unknown mixtures are concerned. However, if a more specific method of analysis is at hand which can be combined with weight, then a small percentage of impurity with a ratio of 1.1 or even 1.05 can be detected.

The reliability of the distribution curve is increased greatly by employing more than one system. Even with closely related isomers or homologs it has, in general, not been difficult to shift β values from 1.0 to 1.5 or more by changing systems. To be sure, optical isomers give β values below the experimental limit and cannot be separated or detected by the method. On the other hand, chemical combination with other optically active solutes give diasterio-isomers which often have more favorable β values.

The reliability of the distribution pattern is also increased by the use of several different methods of analysis for determining a curve. If the compound gives a biological response, quantitative bio-assay should be used. A pure sample gives curves which are superimposable regardless of the method of analysis employed.

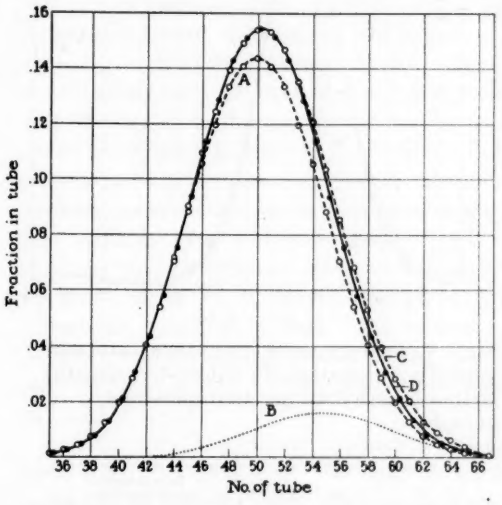


Fig. 3. HYPOTHETICAL 100 TRANSFER DISTRIBUTION of a mixture of 90% compound *A*. ($K = 1.0$), and 10 % compound *B*. ($K = 1.2$). *C* = sum of $A + B$. *D* = a curve calculated for a single substance and matched with *C*.

If satisfactory selectivity is difficult to realize, it is possible to chemically transform the solute to another type which will offer greater selectivity. Valid data are often obtained in this way because of the quantitative nature of the distribution even in spite of inability to cause 100 per cent transformation.

One could continue with generalities indefinitely. However, in the final analysis and in spite of all the theoretical aspects, a method for studying purity must receive prime consideration if it consistently reveals impurity in preparations where other methods have failed. Our studies thus far have been satisfying in this respect. From the very beginning the method proved useful in the antimalarial field (7). Figure 4 is a representative pattern of a good antimalarial preparation; figure 5 of a poor one. Figure 6 (8) is a purity study of a sample of hexanoic acid.

Pentanoic and heptanoic acids or perhaps iso-acids are revealed. Figure 7 is a pattern determined (8) on a mixture of equal parts of myristic, palmitic and stearic acids. The degree of separation and the usefulness of the method in purity studies in this series can be derived from the chart.

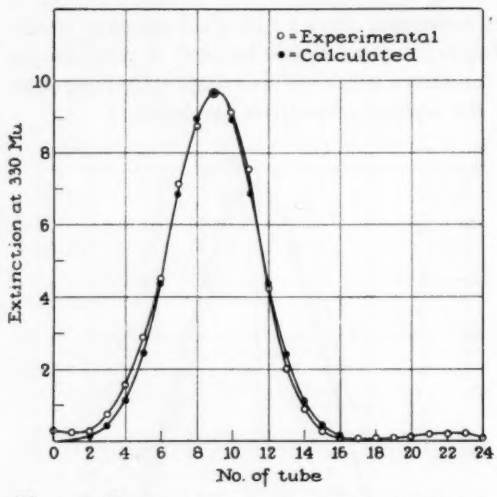


Fig. 4. TWENTY-FOUR TRANSFER DISTRIBUTION of a preparation of 7-chloro-4-(4-dimethylamino-1-methylbutylamino)-quinoline diphosphate.

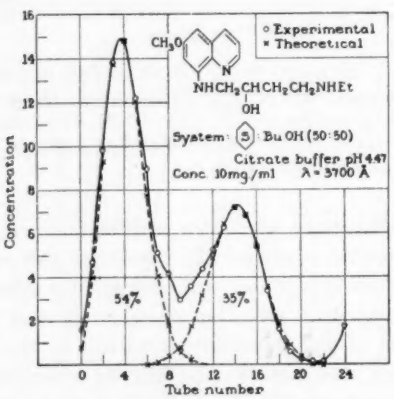


Fig. 5. TWENTY-FOUR TRANSFER DISTRIBUTION of an impure antimalarial.

The penicillin studies are of interest because these substances are known to be stable only under certain conditions. Yet they can be studied by distribution with adequate precision as figures 1 and 2 have shown (5, 9).

Titus and Fried (10) were able to demonstrate the occurrence of more than one streptomycin as shown in figure 8. Their result is of especial interest because of the solubility properties of the streptomycin antibiotics.

Dr. Barry, Dr. Gregory and I have recently attempted the study of antibiotics of polypeptide

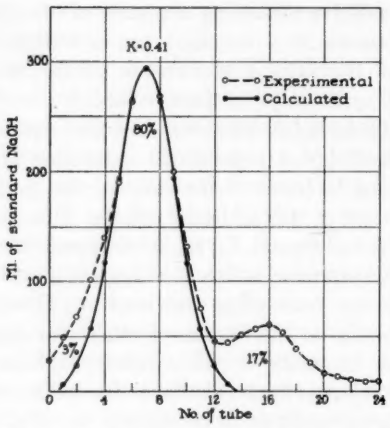


Fig. 6. TWENTY-FOUR TRANSFER DISTRIBUTION of a sample of hexanoic acid.

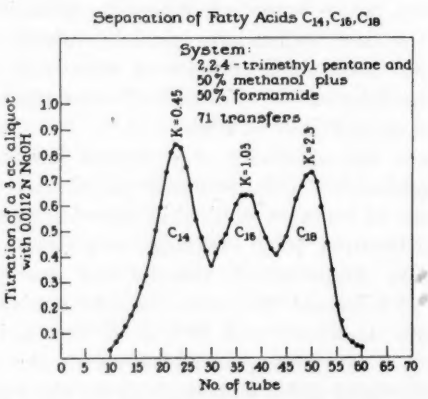


Fig. 7. SEPARATION of higher fatty acids.

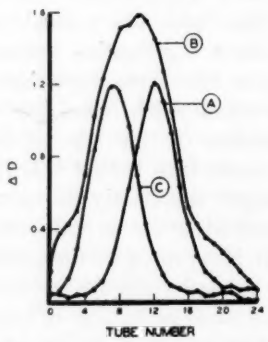


Fig. 8. TWENTY-FOUR TRANSFER DISTRIBUTION of streptomycins.

nature and of higher molecular weight. Significant results have been obtained with the gramicidin-tyrocidine group. The work with bacitracin and

subtilin is less advanced but here too there appears to be definite promise. Figure 9 shows a distribution of a crystalline preparation of gramicidin (11). At least four components are revealed. Three of these have now been isolated and crystallized. Crystalline components of the tyrocidine type have also been obtained. The molecular weight of these substances is believed to be in the neighborhood of 3000.

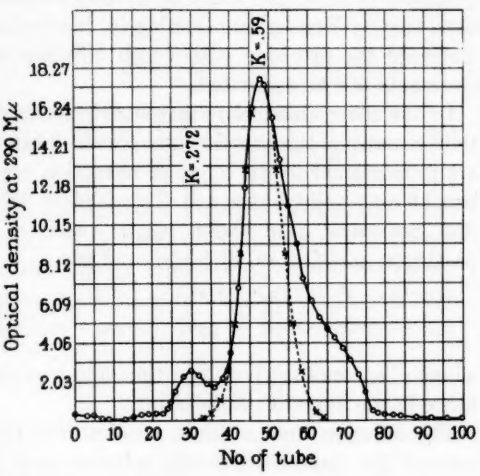


Fig. 9. ONE HUNDRED TRANSFER DISTRIBUTION of crystalline gramicidin. o = experimental; x = calculated.

The work of Woolley in separating polypeptides from the partial hydrolysate of insulin, which is reported at this meeting, further demonstrates the potentialities in this field. Here bio-assay was combined with the method to good advantage.

The combination of the paper chromatographic technique of Consden, Gordon and Martin (12) with counter-current distribution is particularly useful for polypeptides. Each of the fractions from the distribution can be hydrolyzed and analyzed simultaneously for their amino acid spectrum. Shifts in composition are thus plainly revealed with amazingly little work. The starch chromatogram of Synge (13), as improved by the careful work of Moore and Stein (14), would, of course, give the quantitative picture.

Although Martin and Synge (15) considered the process they termed 'Partition Chromatography' to be mainly a liquid-liquid extraction process, a different view is now held by those of us who are closely associated at the Rockefeller Institute. We have had opportunity to study partition chromatography in direct comparison to counter-current distribution which is a true liquid-liquid extraction process. It now appears to us (mainly from the work of Moore and Stein) that the partition chromatograms can be most clearly interpreted in terms of adsorption instead of by liquid-liquid extraction. Such a viewpoint, however, removes nothing from the usefulness of that particular type of chromatography nor from the important contribution made by the English workers. Further, aside from adsorption considerations, counter-current distribution is strictly a discontinuous process operating at complete equilibrium and is thus basically different on another count.

In conclusion, the general applicability of the method may be considered. Obviously it can be applied to all solutes for which a suitable system can be found. Up to the present, systems have been found for a considerable number of bases, fatty acids, the chlorophyls, several complex polypeptides and even substances considered to be only water soluble such as streptomycin. The probability of finding systems for fats and certain sugar derivatives would appear to be good. Perhaps even some proteins are not beyond reach.

REFERENCES

1. EYRING, H. *Anal. Chem.* 20: 98, 1948.
2. CRAIG, L. C. *J. Biol. Chem.* 155: 519, 1944.
3. WILLIAMSON, B. AND L. C. CRAIG, *J. Biol. Chem.* 168: 687, 1947.
4. STENE, S. *Arkiv. Kemi. Mineral. Geol.* 18 A, No. 18, 1944.
5. CRAIG, L. C., G. H. HOGEBOOM, F. H. CARPENTER AND V. DU VIGNEAUD. *J. Biol. Chem.* 168: 665, 1947.
6. VARTERESSIAN, K. A. AND M. R. FENSKE. *Ind. Eng. Chem., Ind. Ed.* 29: 270, 1937.
7. CRAIG, L. C., H. MIGHTON, E. TITUS AND C. GOLUMBIC. *Anal. Chem.* 20: 134, 1948.
8. BARRY, G. T., Y. SATO AND L. C. CRAIG. Unpublished results.
9. BARRY, G. T., Y. SATO AND L. C. CRAIG. *J. Biol. Chem.* 174: 221, 1948.
10. TITUS, E. A. AND J. FRIED. *J. Biol. Chem.* 168: 393, 1947.
11. GREGORY, J. D. AND L. C. CRAIG. *J. Biol. Chem.* 172: 839, 1948.
12. CONSDEN, R., A. H. GORDON AND A. J. P. MARTIN. *Biochem. J.* 38: 224, 1944.
13. SYNGE, R. L. M. *Biochem. J.* 38: 285, 1944.
14. MOORE, S. AND W. H. STEIN. *Ann. N.Y. Acad. Sci.* xlix; 265.
15. MARTIN, A. J. P. AND R. L. M. SYNGE. *Biochem. J.* 35: 1358, 1941.

DIFFUSION, SEDIMENTATION AND ELECTROPHORESIS OF PROTEINS

J. W. WILLIAMS

From the Department of Chemistry, University of Wisconsin, Madison, Wisconsin

Through sedimentation velocity and diffusion studies a large amount of information about the molecular sizes of proteins (and polysaccharides) has been accumulated. In 1940 Svedberg and Pedersen (6) could write that on the basis of such data, "we know that most of the proteins in dilute solutions exist as molecules all having the same well-defined size. For several proteins this molecular size is the same even when the composition of the solution is changed within rather wide limits and is likewise independent of the method of preparing the protein".

For the determination of molecular weight by the usual sedimentation velocity method, independent studies of diffusion on the same solute must be carried out. Since the probable error in the estimated diffusion and sedimentation constants is 2 to 3 per cent, and since the partial specific volume also enters into not completely exact consideration, the molecular weight of a protein determined in this way may be in error by 5 to 10 per cent. Almost another way of saying this is to remark that with the molecular kinetic methods of the 1930's and prior 1940's we may have impurities in our protein preparations of magnitude 10 per cent and still not be able to detect their presence at all. This fact is suggestive of the need for more exact analytical tools as we approach the problem of criteria for purity of biological substances.

DIFFUSION

For the kinetic study of homogeneity or purity of a protein the important measurement is that of diffusion constant. For this purpose it may be used alone or in some combination with sedimentation velocity or electrophoresis. Therefore, we should have before us at least the elementary theory of diffusion as it is used in the study of the physical chemistry of the proteins. The experiments have been made quite accurate in recent years by the development of the scale method of Lamm (5), and the rediscovery and improvement of the schlieren methods of Wiener (10) and of Thovert (8) for the observation of blurring at diffusion boundaries. More recently, the quantitative treatment of the interference phenomena accompanying the deflection of light by the gradients of refractive index in a freely diffusing boundary by Kegeles and Gosting (4) has further increased the precision with which diffusion experiments can be evaluated.

In the more classical refractometric methods, the ordinates of the experimental curves obtained are proportional to the index of refraction gradient at each point in the cell. To calculate diffusion constants from these, use is made of a particular solution of Fick's second law,

$$\frac{\partial c}{\partial t} = \frac{D\partial^2 c}{\partial x^2} \qquad [1]$$

where c = concentration, t = time and x = distance in the direction of diffusion.

For diffusion from a dilute solution into pure solvent the desired particular solution may be written as

$$\frac{dn}{dx} = \frac{n_1 - n_0}{2\sqrt{\pi Dt}} e^{-x^2/4Dt} \qquad [2]$$

where dn/dx = index of refraction gradient, n_1 = index of refraction of solution, n_0 = index of refraction of solvent and D = diffusion constant.

In the scale method, the experimental curve is a plot of the displacement of the lines of a uniform scale photographed through the system in which diffusion is taking place as a function of the distance of these lines from an arbitrary reference line. For the calculation, the best curve is drawn through these points and traced onto another sheet with an arbitrary origin chosen for convenience. Scales of abscissae and ordinates are chosen and a table of ordinates (S) for equal intervals on the axis of abscissae (s) constructed. Since the origin is located by estimating the centroid in tracing the curve, a generalized form of *equation 2* is used, which takes into account the distance of the true centroid from the origin

$$S = \frac{N\omega}{\sigma\sqrt{2\pi}} e^{-(s-\beta)^2\omega^2/2\sigma^2} \qquad [3]$$

where S = scale-line displacement corresponding to scale-line distance s from chosen origin, σ^2 =

$2Dt$, β = distance of chosen origin from centroid, ω^2 = numerical constant and $N\omega$ = area of curve.

If this analytical expression is the equation for the experimental curve, values of D may be calculated from the s, S table by using the properties of the curve. The two most frequently used equations for D are based on the height and area (4a) and the second moment and area of the curve (4b).

$$D = \frac{(\Sigma S)^2}{S^2_{\max.}} \frac{\omega^2}{2\pi t} \qquad [4a]$$

$$D = \left[\frac{\Sigma s^2 S}{\Sigma s} - \left(\frac{\Sigma s S}{\Sigma S}\right)^2\right] \frac{\omega^2}{2t} \qquad [4b]$$

Since the units of s and S in *equation 3* are arbitrary, it is necessary to convert the values from different experiments to the same (dimensionless) units in order to be able to compare them. This is usually done by the transformations

$$\Xi = \left[\frac{5\sigma}{\omega \Sigma S}\right] S$$
$$\xi = \frac{(s - \beta)\omega}{2\sigma} \qquad [5]$$

The resulting set of values of Ξ and ξ may then be compared with those from other experiments and with the corresponding normal (Gauss) curve.

$$\Xi = \frac{5}{\sqrt{2\pi}} e^{-2\xi^2} \qquad [6]$$

Coincidence of the experimental points with the normal curve shows that *equation 2* applies to the diffusion system studied.

The assumption that *equation 2* does describe the variation of index of refraction gradient (and so of concentration gradient) as a function of distance in the diffusion cell and of time, so long as diffusion has not progressed far enough to affect concentrations at the end of the cell, has been justified by a great mass of experimental data. As a result, coincidence of the experimental with the normal curve is taken as an indication of homogeneity, or, more strictly and properly, no inference about the solute is drawn from coincidence, but deviations of the experimental curve from the normal are attributed to some peculiarity of the solute, including the presence of impurity. Thus, this method is not entirely satisfactory.

Equation 2 may be solved for D to give

$$D = \frac{(n_1 - n_0)^2}{\left(\frac{dn}{dx}\right)^2_{\max.} 4\pi t} \qquad [7]$$

since dn/dx has its maximum value at the center of the peak where $x = 0$. In the Kegeles-Gosting interference method the determination of the diffusion constant is based on measurements of the positions of interference bands formed by the schlieren lens and the diffusion gradient in the plane of the knife edge. The important formula is

$$D = \frac{(n_1 - n_0)^2}{\left(\frac{C_t}{ab}\right)^2 4\pi t} \qquad [8]$$

where $C_t = \frac{Y_j}{e^{-z_j^2}}$ and Y_j is the downward displacement of j'th interference fringe from the undeviated slit image in the plane of the knife edge. The quantity $e^{-z_j^2}$ is obtained from tables and is a function of the number of the particular fringe and the total number of fringes in the interference pattern. Thus C_t becomes the maximum downward displacement of light in the focal plane of the schlieren lens predicted by geometrical optics and

$$\frac{C_t}{ab} = \left(\frac{dn}{dx}\right)_{\max} \qquad [9]$$

The downward displacement Y_j below the undeviated slit image in terms of diffusion constant is

$$Y_j = \frac{ab(n_1 - n_0)}{2\sqrt{\pi Dt}} e^{-z_j^2} \qquad [10]$$

A value of D from each fringe may be obtained by an accurate measurement of the values of Y_j. Since Y_j is usually of the magnitude 10 mm. and may be measured to within a few microns, this method of measuring the diffusion constant of an ideally diffusing substance should be capable of another order of magnitude of accuracy as compared to that obtained from the scale line displacement methods.

Actually this is true, and as a result we are provided with a more sensitive method for the detection of the abnormalities in diffusion which are caused by the presence of impurities in a protein solute (or by concentration dependence of diffusion). In the classical Lamm scheme indication of impurity is the failure of the normalization, which as suggested above is not sensitive.

In the interference method indication of impurity is lack of constancy of the maximum downward displacement C_t, as computed from successive fringes. The values $(n_1 - n_0)$ must be measured in an independent experiment.

It will be well to emphasize that there has been no gain in theoretical precision. The beneficial result comes from the fact that it is possible to measure the fringe displacements with one order of accuracy higher than is possible in the case of the scale-line displacements. Detailed consideration of this general problem will be the subject of a subsequent scientific contribution.

We should mention that Coulson, Cox, Ogston and Philpot (3) have just published a description of a rapid method for determining diffusion constants in solution which is based on the interference principle. The basic theory is the same as that of Kegeles and Gosting, but the English investigators have not as yet developed the theory to the point where the detection of the presence of impurities has become possible.

SEDIMENTATION VELOCITY

The usual sedimentation velocity criterion for the homogeneity of a protein is that a peak in the scale line displacement-distance curves moves in the centrifugal field as a single apparently symmetrical boundary when the protein is dissolved in buffers of various hydrogen ion concentrations and ionic strengths. Actually, this is judgment based largely upon diffusion and, to be more precise, the investigation should be made to determine whether the boundary has been broadened solely by diffusion. This may be accomplished by comparing the diffusion curves obtained in the ultracentrifuge experiments (corrected for the sector shape of the cell) with the theoretical curves calculated from values of D found by independent diffusion experiments. If the two sets of curves coincide we have a good criterion of monodispersity—in this instance all the molecules have the same specific sedimentation constants. Unfortunately, this test is hardly to be recommended because it is difficult to maintain temperatures of sufficient constancy in a centrifuge cell so that convective disturbances can be avoided. Slight convective inhomogeneity strongly influences the apparent diffusion constant and failure of actual and theoretical curves to coincide might be misinterpreted.

Again, if the molecules are large (low diffusion constants) differences in sedimentation constant as small as a few per cent may be readily detected in the sedimentation diagram. When, however, the molecules are small (high diffusion constants), two substances having even as high as 25 per cent difference in sedimentation constant may appear from the sedimentation diagram to be a uniform substance with sedimentation constant intermediate between that of the single components. Calculation of the apparent diffusion constant will indicate the nonhomogeneity by giving values which show a pronounced drift with time, and the average constant obtained in this way will be considerably higher than that found from special diffusion experiments.

ELECTROPHORESIS

Diffusion and sedimentation experiments may give information about the mass (and shape) homogeneity of proteins. A very satisfactory physical chemical test for another kind of homogeneity of a protein is found in the electrophoresis experiment of the present day. In it a quantitative measure of the charge heterogeneity may be obtained through the determination of the standard deviation of the mobility distribution, assuming that the protein has a Gaussian distribution of mobilities.

The usual criterion for the electrophoretic homogeneity of a protein is that it migrates as a single boundary in an electrical field in buffers of various hydrogen ion concentrations and ionic strengths. To this criterion must be added a second one to the effect that the rate of spreading of the protein boundary under conditions such that convection and anomalous electrical effects are avoided should be no greater than that due to diffusion alone. If this spreading is greater than can be accounted for on the basis of diffusion and the gradient sharpens on reversal of the current, the protein may be heterogeneous with respect to electrophoretic mobility even if a single symmetrical boundary is observed.

Tiselius pointed out that the reversible spreading due to lack of homogeneity of the migrating substance with regard to mobility may be distinguished from that caused by the boundary anomalies because the latter effects are always of opposite sign at the two boundaries, while in the case of inhomogeneous proteins both boundaries behave in the same manner. Tiselius and Horsfall (9), and Sharp, Taylor, Beard and Beard (7) have proposed quantitative methods for the representation of boundary spreading, but neither

group gave sufficient attention to the actual distribution in mobility, even in one variety of protein molecules and to the superimposed diffusion.

There are four factors which influence the rate at which a protein gradient will spread under the influence of the electrical field. Two of them are irreversible, two reversible. 1) Diffusion is superimposed irreversibly on electrophoresis. 2) Convections due to temperature gradients set up in the cell by electrical heating produce irreversible changes. 3) Conductivity and pH differences across the boundaries give reversible changes (which can be largely eliminated by conducting the experiment at the isoelectric point). 4) Actual electrochemical inhomogeneity of the protein molecules gives rise to reversible electrophoretic spreading (broadening of the peak due to differences in net charge or size or shape).

Sharp and collaborators (7) have shown that when diffusion during an electrophoresis spreading experiment is negligible compared to the electrical spreading, the mobility distribution may be obtained from the refractive-index gradient curves, and a heterogeneity constant (H) may be calculated from the time rate of change of the standard deviation of the gradient, $\frac{\Delta\sigma}{\Delta t}$.

They write

$$H = \frac{\Delta\sigma}{\Delta t E}, \qquad [11]$$

where E is the electric field strength.

In the case in which diffusion of the protein during the experiment is not negligible and the mobility distribution may be represented by the Gaussian probability function, Alberty (1) has shown that the experimental refractive-index gradient at the isoelectric point will have Gaussian form, and a heterogeneity constant, h, may be calculated by using *equation 12*.

$$D^* = \frac{\sigma^2 - \sigma_0^2}{2t_E} = D + \frac{E^2 h^2}{2} t_E \qquad [12]$$

In this equation σ_0 is the standard deviation of the gradient curve at the moment the field is applied, and σ is the standard deviation after electrophoresis for t_E seconds. According to this equation, the 'apparent' diffusion constant, D^*, should plot as a straight line against time of electrophoresis and extrapolate back to the normal diffusion constant, D, at zero time. In order to apply this method it is necessary that all the protein molecules in the sample have the same diffusion constant. Also, the protein must be soluble and stable at its isoelectric point. These conditions are fulfilled for the proteins studied here in Madison. The heterogeneity constant, h, is actually the standard deviation for the mobility distribution $g(u)$.

$$g(u) = \frac{1}{h\sqrt{2\pi}} e^{-u^2/2h^2} \qquad [13]$$

The mobility distribution for the protein may therefore be plotted by using the value of h determined from the slope of the graph of D^* vs. t_E and tabulated values for the Gaussian probability function.

As a check on the elimination of convection, the field should be reversed for an equal period of time to bring the boundary back to its initial state except for diffusion. If the field is reversed at time t_1, the apparent diffusion constant during the reversal period is given by *equation 14*.

$$D^* = D + \frac{E^2 h^2}{2} \frac{(2t_1 - t_E)^2}{t_E} \qquad [14]$$

According to this equation the apparent diffusion constant becomes equal to the true diffusion constant at $t_E = 2t_1$. *Equation 14* may also be used to calculate the heterogeneity constant from data obtained during the reversal period.

A protein which shows reversible spreading in electrophoresis at the average isoelectric point must contain molecules with isoelectric points higher than the average and molecules with isoelectric points lower than the average. The isoelectric point distribution may be determined electrophoretically by studying the reversible spreading at several pH's in the isoelectric range. If the rate of change of the average mobility of the protein with pH, $du/dp\text{H}$, is constant in this range and the heterogeneity constant is independent of pH, the mobility distribution may be used directly to calculate the isoelectric point distribution. For instance, these conditions appear to apply very well in the cases of human γ_2-globulin and horse pseudoglobulin. The isoelectric point, pI, of molecules with a mobility u at the average isoelectric point of the protein pI_{av} is given by *equation 15*.

$$pI = pI_{av} - \frac{u}{\frac{du}{dpH}} \qquad [15]$$

Detailed electrophoresis spreading experiments have been carried out with human γ_2-globulin,

immune lactoglobulin, bovine serum albumni, β-lactoglobulin, ovalbumin, ribonuclease, and a number of others. (2) All of these materials had been carefully prepared by recognized methods and in some cases were crystalline in form. With the possible exception of ribonuclease none of the proteins were found to be completely homogeneous. Of course, it is not surprising that the antibody proteins among them are not homogeneous, since they contain many different antibodies.

CONCLUSION

In this report we have emphasized that a symmetrical stationary boundary in diffusion or moving boundary in sedimentation or electrophoresis is not sufficient evidence for the homogeneity of molecularly dissolved proteins and that careful attention must be devoted to the exact mathematical form of the curve which represents the boundary. We have selected as the basis for discussion two new approaches which have been developed in this Laboratory, the Kegeles-Gosting interference method for the study of diffusion and the Alberty treatment of reversible boundary spreading in electrophoresis, as being steps in the improvement of the more classical kinetic methods so that detection of heterogeneity in protein systems becomes more satisfactory.

It has been our experience that a number of proteins obtained from several laboratories in the United States, reputedly of crystalline form and constant solubility, cannot now meet these new kinetic tests for homogeneity. Actually, the interference method for diffusion is so sensitive that momentarily we have put aside work with it as an instrument for the characterization of proteins and have gone back to some of the purified simple sugars, simple electrolytes and amino acids for study, expecting that as the method is further developed and consolidated preparative procedures for proteins will have been improved to provide more satisfactory solutes. The information which has been here collected is intended to be of aid in biochemistry.

REFERENCES

1. Alberty, R. A. *J. Am. Chem. Soc.* 70: 1675, 1948.
2. Alberty, R. A., E. A. Anderson and J. W. Williams. *J. Phys. and Colloid Chem.* 52: 217, 1948; Alberty, R. A. and E. A. Anderson. Forthcoming publication.
3. Coulson, C. A., J. T. Cox, A. G. Ogston and J. St. L. Philpot. *Proc. Roy. Soc.* A192: 1382, 1948.
4. Kegeles, G. and L. J. Gosting. *J. Am. Chem. Soc.* 69: 2516, 1947.
5. Lamm, O. *Nova Acta Soc. Sci.* Upsala 4 (10): 1937.
6. Svedberg, T. and K. O. Pedersen. *The Ultracentrifuge.* New York: Oxford University Press, 1940.
7. Sharp, D. G., A. R. Taylor, D. Beard and J. W. Beard. *J. Biol. Chem.* 142: 193, 1942.
8. Thovert, J. *Ann. Phys.* (9) 2: 369, 1914.
9. Tiselius, A. and F. L. Horsfall. *Ark. Kem. Min. Geol.* 13A, No. 18, 1939.
10. Wiener, O. *Wied. Ann.* 49: 105, 1893.

SOLUBILITY METHOD OF ANALYSIS

ROGER M. HERRIOTT

From the Rockefeller Institute for Medical Research, Princeton, New Jersey

The solubility method as a test of purity has been examined in very few instances (1–3) with mixtures of two or more known components. In spite of this deficiency, the method has the following general properties to recommend it: (*a*) It is based on the thermodynamic principles of heterogeneous equilibria which are among the soundest of theoretical concepts in chemistry. The phase rule of Gibbs was derived by purely mathematical reasoning and is not dependent on any assumption with regard to theories of kinetics or structure of matter. (*b*) The method is applicable to all species of molecules. (*c*) The method is sufficiently sensitive to distinguish between *d* and *l* optical isomers although as an unknown a 1:1 mixture of isomers would analyze as a simple component. With some knowledge of the nature of the unknown special solubility experiments would reveal the two-component nature of even this system. (*d*) The equipment necessary is very simple and is probably present in every laboratory. (*e*) The quantity of material may be small for the limiting factor is usually the method of analysis. (*f*) If the material under test is impure, the solubility diagram indicates a method of separating the components, and finally; (*g*) Solubility experiments can be designed which with a minimum of effort will answer otherwise difficult problems such as the identity of synthetic and natural occurring products or distinguish between a mixture of optical isomers and a racemic compound.

On the debit side there are a few systems of two components which as an unknown would analyze as a single component. Examples of these are a 1:1 mixture of *d* and *l* optical isomers, and Webb (4) has pointed out that isotopic species will be indistinguishable by this method.

Details of the theory and practice of the solubility method have been amply discussed (1, 5, 9–11), so that only those aspects which are necessary to the discussion or need emphasizing will be considered.

METHOD

Briefly the solubility method consists of *a*) mixing various quantities of material under investigation with equal or known volumes of solvent until equilibrium is attained; *b*) separation of the solid phase from the solutions; *c*) determination of the concentration of the material dissolved in the various aliquots; and *d*) plotting the concentration of dissolved material against the total (solid and dissolved) per unit volume.

INTERPRETATION OF RESULTS

A plotting of the results of the solubility data will take the form of one of the three general type curves shown in figure 1.

In those aliquots containing no solid phase at equilibrium, the dissolved concentration equals the total added and therefore the points fall on the *line OA*. This line bears at an angle of 45°C. from either coördinate when the scale units are equal. As the total concentration is increased a point will be reached where a small amount of solid will remain after equilibration. The manner in which the concentration of dissolved material varies with increasing total material beyond that point where the first solid phase persists is the crux of the solubility method. The solubility may vary in any of three ways. 1) It may remain constant as represented in figure 1 by *curve OAE*. Such a curve represents *a constant solubility* and is obtained when the material is *a*) a single component (i.e., pure); *b*) a solid solution of two or more materials having identical solubilities; and *c*) a mixture of two or more materials which are present in the preparation in direct proportion to their solubilities. Cases *b* and *c* are not often encountered and may be detected by performing the solubility determination in solvents of a different nature or possibly by varying the temperature. In general the ratio of the solubilities of materials varies with temperature and the nature of the solvent. If the material under study has a constant solubility in several different solvents, it is probably a single component. A racemic 1:1 mixture of *d* and *l* isomers or a mixture of isotopic isomers would behave as a single component even in several solvents. With information as to the nature of the material the optically active isomers could in some instances be resolved by the classi-

cal method based on the difference in solubility of the complex salts formed with an optically active reagent. Use of a solution saturated with respect to one of the antipods as solvent would also bring out the two-component nature of even the 1:1 mixture. In this case the solid phase at equilibrium would be rich in that antipod used to saturate the solvent.

2) The solubility may increase linearly as the total increases and change abruptly to a different linear positive slope or to one of zero. This type result is obtained when the components form simple mixtures in the solid and their solubilities are independent of one another. Each time there is an abrupt change in slope a new solid phase appears. *Curve OBCE* represents such a system with pure solid phases appearing at *B* and *C*. A mixture of *d* and *l* optical isomers in any proportion other than 1:1 would be expected to have a curve of this type.

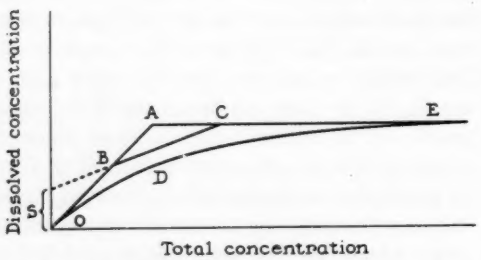


Fig. 1. GENERAL TYPES of solubility curves.

Northrop and Kunitz (5, 1) have shown that the value of the intercept of the *line BC* with the ordinate is the solubility *S* of the pure component which first appears as a solid phase. The quantity of this pure component, as a fractional part of the total, is given by one minus the slope of the *line BC*. The corresponding values of the second component are obtained by difference. Kunitz and Northrop (1) obtained gratifying confirmation of their predictions upon mixing relatively pure preparations of two proteins. More recently Thorp (3) has similarly found mixtures of known isomers of *DDT* to respond quantitatively to this method of analysis.

(3) Third and finally the solubility may vary continuously as the total is increased (see *curve ODE*) approaching asymptotically a constant value. Such a result is obtained with solid solutions. In such cases no quantitative predictions are possible. If one assumes that Raoult's law holds even qualitatively, it may be predicted that the first *precipitate* to appear along th curve is richer than the starting material wit respect to the more insoluble component. Sim larly the *solution* in the presence of an excess solid is richer with respect to the more solub component.

The behavior of a solid solution is similar t that of a system in which one of the solute com ponents is a liquid of limited solubility but i which the second solute component is solubl such a system conforms to the laws of partitio or distribution.

Solubility method and the phase rule. The phas rule is a useful generalization of heterogeneou equilibria relating the number of component phases and degrees of freedom. Use has bee made of it to establish the solubility method o sound theoretical principles. However, it is no necessary to apply the phase rule or even unde stand it to make considerable use of the sol bility method. Most of the work with th method is for the purpose of deciding whethe a preparation consists of one or more component The simple concept that an equilibrium conditio is independent of the quantities of any of th phases will cover most of such cases, for if the sol bility value is independent of the quantity solid phase, it is *a constant solubility*, and in ge eral that is the characteristic of a single componen

It may be worth pointing out that the phas rule is not concerned with the time required to ge to equilibrium, nor does it indicate that the solut in solution is identical with that in the soli phase. In many instances they are known not t be identical.

A reversibly dissociable solute does not presen a different problem in the solubility method. I such a system there is only one more componen even if the solute dissociates into two parts i the concentration of the undissociated and on part will define the concentration of the third With this one more component there will be on more degree of freedom. The concentration o the dissociated components is usually fixed by th concentration of undissociated material and th particular solvent used. Thus a dissociation o any similar reversible phenomenon does no jeopardize the usefulness of the solubility metho by limiting it to nondissociating systems or re quiring that the existence of such dissociation b known. If the dissociation is not reversible, the an equilibrium cannot be established and th method is not applicable.

For those interested in a phase rule proof of e method, the following is a brief description. is rule in its abbreviated form is $P + F = + 2$ where P = the number of phases (physi- lly separable portions), F = the number of egrees of freedom (usually temperature, pres- re and concentration of components) and C = e number of components (the smallest number independent constituents which will define the stem).

In solubility experiments the temperature and essure, two degrees of freedom, are held con- nt so the equation simplifies to $P + F' = C$ here $F' = F - 2$. Table 1 shows how these fac-

TABLE 1

SYSTEM	REGION OF CURVES IN FIG. 1	NUMBER OF			
		P	F	C	
				solvent	solutes
olution (no solid)	OB	1	1 or more	1	1 or more
ingle pure solid in a saturated solution	AE	2	0	1	1
ingle pure solid in solution of two components	BC	2	1	1	2
ixture of two pure solids in a saturated solution	CE	3	0	1	2
olid solution and corresponding solvent solutions	ODE	2	1 or more	1	2 or more

rs vary in the different solubility diagrams of gure 1.

QUANTITATIVE LIMITATIONS OF THE METHOD

The limit of resolving power of this method is pendent on *a*) the relationship of the com- nents in the solid phase, *b*) the solubilities of e individual components and *c*) the precision the analytical method. If the solid phase is a lid solution, no quantitative analysis of the rve is possible. In general the greater the differ- ce in solubility of the individual components, e greater will be the resolution. No one analyti- l method can be used with all materials but the dry weight of solute has a very wide range of usefulness and can be a very precise measurement. Other properties must be measured when volatile solutes are under study or when concentrated salt solutions are employed as solvents. Analyses of several physical and chemical properties, specific and nonspecific, can often be rewarding, especially when examining an unknown material having specific biological activity.

The recent publication by Webb (3) that 10 per cent *l*-leucine in the presence of 90 per cent *dl*-leucine resulted in a slope of only 1 per cent serves as a warning against complete reliance being placed on the method. It must follow from this result of Webb that this system is not a simple mixture of two components. Either there is solid solution formed or an additional component. In general the solubility method is open to the same difficulties and criticisms as the melting point determination to which the solubility method is related. These difficulties, however, have not prevented the melting point determination from being an extremely useful criterion of purity.

ADDITIVE NATURE OF SOLUBILITY AND ITS USE IN SPECIAL CASES

Special experiments based on the fact that in general solubility is an additive property have been used with marked success in answering problems concerning the identity of two preparations. Even closely related proteins or *d* and *l* optical isomers can be distinguished in this way.

Landsteiner and Heidelberger (6) showed that hemoglobins from closely related species were not identical since the solubility of the two together was greater than either by itself. Similarly Northrop (7) found that whereas the enzymatic activity, crystalline form, optical rotation, diffusion constant, and even the solubility of crystalline bovine and swine pepsins were not significantly different, a solution saturated with swine pepsin dissolved bovine pepsin. Similar procedures were used by Loring and du Vigneaud (2) to show that a preparation of 'stone' cystine was the same as hair cystine and that a racemic mixture of *d* and *l* cystine was easily distinguished from the racemic compound. The writer (8) has recently made use of this principle to show the probable identity of synthetic *dl*-3 monoiodotyrosine and a product isolated from iodinated pepsin.

SEPARATION OF COMPONENTS BY METHODS BASED ON THE SOLUBILITY DIAGRAM

It follows from the discussion in the section on 'Interpretation of Results' that if the system under examination is shown by the solubility method to consist of two or more components, that a procedure for their separation based on this diagram can be developed. Thus if the components are simple mixtures (case II), the first solid phase to appear, represented in figure 1 by the region enclosed in the *triangle ABC*, will be a pure component having the solubility *S* indicated in figure 1. If the components form solid solutions (case III) and their behavior does not deviate qualitatively from that predicted by Raoult's law, the first *solid phase* to persist at equilibrium will be richer in the least soluble component than the starting material while the *solution* in equilibrium with an excess of solid phase will be richer with respect to the most soluble component.

Diphtheria antitoxin (12) and swine pepsin (13) preparations whose solubility curves showed a high degree of inhomogeneity were purified by procedures based on the last of the above-mentioned suggestions. The procedures consisted of extracting the materials under conditions which dissolved only a fraction of the total, saving the solution and, after precipitating out the protein, repeating the extraction procedure on this last precipitate. With both proteins the solubility curves improved (i.e., approached more nearly that of a single component) with each extraction, and became indistinguishable from that of a single component after the third extraction.

GENERAL PRECAUTIONS

The solvent and conditions should be chosen so there is a minimum destruction or permanent alteration of the solute during the examination. A mixed or split solvent (two miscible liquids) is often useful in adjusting the solubility of the solute to a value which fits the needs of the particular experiment.

In general a state of equilibrium may be considered as established if approaching the same conditions from the under saturated and supersaturated sides results in the same solubility. In each experiment the solubility value must be determined at two intervals of time for the time of equilibration cannot be predicted with certainty. In those tubes containing very little solid phase more time is required to saturate than in those containing a large excess. If too long an equilibration is permitted secondary reactions *may* develop. In this connection the experiments of Loring and du Vigneaud (2) are of interest. In their work a mixture of *d* and *l* cystines dissolved in water forming a solution saturated to these components, and the solubility value remained unchanged for 10 to 20 hours. After this time the solubility decreased to about a quarter of the previous value. This low value agreed with that found for the racemic compound *dl* cystine. The authors also stated that microscopic examination of the solid phase showed that it was all converted to the racemic compound.

Solutions saturated with amorphous protein are of course supersaturated with respect to the crystals and if the latter begin to form the solubility will decrease. These examples merely serve to illustrate how a true equilibrium may exist for only a relatively short time.

SUMMARY

The solubility method is a theoretically sound, highly selective method of detecting the presence of impurities. It may be used with any kind of molecule and in general can be made as accurate as the method of analysis. No special apparatus is required. When the results indicate the presence of a second component, it is possible to separate one and in some instances two components into single component fractions.

REFERENCES

1. Kunitz, M., and J. H. Northrop. *Symposia on Quan. Biol. Cold Spring Harbor*, 6: 325, 1938.
2. Loring, H. S., and V. du Vigneaud. *J. Biol. Chem.* 107: 267, 1934.
3. Thorp, D. *J. Soc. Chem. Ind.* 65: 414, 1946.
4. Webb, T. J. *Anal. Chem.* 20: 100, 1948.
5. Northrop, J. H. and M. Kunitz. *J. Gen. Physiol.* 13: 781, 1930.
6. Landsteiner, K. and M. Heidelberger *J. Gen. Physiol.* 6: 131, 1923.
7. Northrop, J. H. *J. Gen. Physiol.* 16: 615, 1933.
8. Herriott, R. M. *J. Gen. Physiol.* 31: 19, 1947.
9. Butler, J. A. V. *J. Gen. Physiol.* 24: 189, 1940.
10. Herriott, R. M. *Chem. Rev.* 30: 413, 1942.
11. Northrop, J. H., M. Kunitz and R. M. Herriott. *Crystalline Enzymes.* (2nd ed.) New York: Columbia University Press, 1948.
12. Northrop, J. H. *J. Gen. Physiol.* 25: 465, 1941.
13. Herriott, R. M., V. Desreux and J. H. Northrop. *J. Gen. Physiol.* 24: 213, 1940.

Symposium on Hemin Pigments and Chromoproteins[1]

DAVID L. DRABKIN, *Chairman*

DISTRIBUTION AND METABOLIC ASPECTS OF DERIVATIVES OF IRON PROTOPORPHYRIN (HEMIN)

DAVID L. DRABKIN

Department of Physiological Chemistry, Graduate School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania

Hemoglobin, as a prototype of naturally occurring pigments and the mammalian body's most abundant specialized chemical, is a complex of porphyrin, iron and protein. Appropriately to its three-fold character the investigative attack upon this compound and its natural or laboratory relatives has been multiple. Except to disclose details, resort to artifice has been unnecessary to 'label' hemoglobin, endowed with a distinctive flag, iron protoporphyrin, which, at half-mast or full, can best be seen with a spectroscope. But, this is by no means the only applicable physical probe, for hemoglobin has been studied with a greater variety of methods than any other protein. Gasometry (1), potentiometry (2), spectrophotometry (3), measurement of paramagnetic susceptibility (4), behavior in a gravitational field (5) and classical (6) as well as x-ray crystallography (7) have contributed their share to present total information.

Step by step the organic structure has been elucidated (8)—of the porphyrins, of the metalloporphyrins, and of the latter combined with nitrogenous bases—until a near synthesis or reconstitution of the grand whole, hemoglobin, has been accomplished, when 'native' globin was available (9). The proteins of the complexes have been dissected into their constituent amino acids (10); their relatively basic character established. Yet, the protein of hemoglobin has jealously guarded two conjoined secrets—how it protects the iron from rusting, and how it participates in the union with oxygen. Among complexes of iron, ferrohemoglobin remains unique; *its iron does not rust on exposure to molecular oxygen and moisture*, but, without change in valence, combines stoicheiometrically and reversibly with the gas in a process of oxygenation-deoxygenation, the basis of its biological function. In contrast, ferrocytochrome *c* oxidase is *autoxidized to the ferric state* (and at low pressures of molecular oxygen). In this process oxygen is left as a carcass, the oxygen ion, eventually to be carted off, and the aerobic spark is supplied to initiate the utilization of the energy inherent in metabolites. Thus, it is remarkable that nature has assigned to derivatives of iron porphyrin the major rôles of both transport and utilization of oxygen, though the know-how of oxygenation is restricted to the hemoglobins.

Cells need energy. In this modern age the slogan 'need for oxygen' has given way to the more fundamental 'need for energy.' Oxygen functions to 'spark-plug' the energy-yielding processes in aerobic cells, and the 'respiratory' chemicals, hemoglobin, myoglobin, cytochrome *c*, etc., are only agents in a broader process which cuts through every segment of biological organization. An attempt has been made in figure 1 to depict in outline the biological plan. Certain features may be pointed out. *a*) There are two main integrated areas of oxygen transport and of oxygen utilization. The former is poised to resist oxidation, the latter to favor it. *b*) Regardless of whether biological enzyme systems may be made to work in a cell-free medium, *in vivo* we deal with organized units, the cells, whose membrane barriers offer advantages and impose limitations, which at times may be striking. *c*) Some of the hazards, besides cytolytic agents, to which these biological processes may be exposed are illustrated in the figure by an indication of the locus of action of various histotoxic agents, CO, CN^-, F^-, iodoacetate, barbiturates and certain oxidants. To be effective, these substances must be able to penetrate the cellular membranes. *d*) More than 90 per cent of the effective free energy for cellular work is derived from the process,

[1] Atlantic City, N. J., March 1948.

whose overall accomplishment is the oxidation of hydrogen in metabolites, yielding H_2O. The relationship of respiratory CO_2 to O_2 is no longer, as it was for over a century, the simple matter of the oxidation of carbon. Certain aspects of the various processes will be discussed.

Distribution and natural occurrence. We have been gathering information on the total quantities of the chromoproteins in different species (11, 12). Table 1 is illustrative of the results.

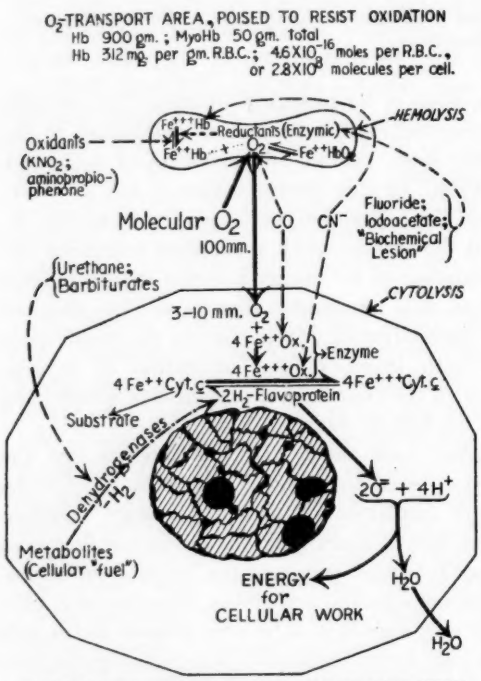


Fig. 1. SCHEMA OF THE BIOLOGICAL PLAN of integration of the areas of oxygen transport and oxygen utilization. The red blood corpuscle, *above*, and liver cell, *below*, are drawn to represent relative size.

The values for the quantity of hemoglobin in the species studied appear to be more or less directly proportional to the body mass. Myoglobin is particularly high in the muscles of the horse, a hard-working animal, and low in the muscles of man. Comment is unnecessary, except that this fits the concept of myoglobin as a chemical oxygen storer, or transporter "in time instead of space" (13).

The most interesting findings are for total cytochrome *c*, involved in the utilization of oxygen. The quantitative estimation of this pigment was by means of a direct microspectrophotometric technique, developed in our laboratory (14). If rat and man only are considered, a very good case is made out for the identity of the cytochrome *c* values, when expressed on the basis of 1 m^2 of surface area, or of a fractional exponent of the body weight. We shall not quibble concerning academic arguments (15) as to which of the above bases of reference is the more correct, for insistence upon this may steer us away from a possibly significant, fundamental relationship. On the other hand, the horse is exceptional. This species has too much cytochrome *c*, and, thereby damages the hope for a generalization. However, the exception may be instructive. There is a philosophical exit from this dilemma, which may be worth considering. It involves, among other matters, an extension of the concept of 'basal metabolism', which unfortunately has carried with it the idea of 'fixity'. The utilization of oxygen and the need for cytochrome *c* function will increase with work, and the body content of cytochrome *c* may be related to a 'working metabolism', or to a metabolic 'capacity', which expresses the degree of expansibility of metabolism (probably under complicated hormonal as well as other control) from a basal level to a severe work level. In the horse the working metabolism probably reaches higher levels than in the other species. With this in view we were led to look into the cytochrome *c* content of the placid cow in contrast with the excitable, hard-working horse. While our data on the bovine species are limited, the cow fits the generalization of proportionality of the cytochrome *c* content to a fractional exponent of the body weight. In view of the findings, I believe that in rat, man and cow, the ratio of the working to the basal metabolism must be the same. This metabolic relationship of cytochrome *c*, though imperfect, appears to be one of the most striking, thus far uncovered, for a cellular chemical component.

Also of interest are the relative quantities of the three chromoproteins: In a 250-gm. rat, hemoglobin : myoglobin : cytochrome *c* = 222:7:1; in a 70-kgm. man and in a 500 kg. horse, the respective ratios are 1150:51:1 and 359:112:1 (from data in table 1). Quantitatively, the predominant position of hemoglobin, the oxygen transport chemical, is apparent. The transportation system appears to be as important biologically in mammals as it is in modern, social organization. Not obvious in the above values,

however, is the concentration per functioning cell. In both rat and man, the concentration of hemoglobin per gm. of erythrocytes is about 312 mg. (16) (fig. 1). The larger liver cells contain, respectively, approximately only 0.2 and 0.02 mg. of cytochrome *c* per gm. When these very discrepant values are reduced to molecular magnitudes, i.e., either to moles per cell or to molecules per cell (17), a rather amazing result is obtained: Hemoglobin in rat and man = 2.8×10^8 molecules per cell; cytochrome *c* in rat = 4.2×10^7 molecules per cell (a value of the same magnitude as that of hemoglobin); cytochrome *c* in man = 4.2×10^6 molecules per cell. Operating in this, at first sight, puzzling situation are the simple factors of relative molecular and cell sizes: Hemoglobin is 5 times larger than cytochrome *c* and the liver cell some 47 times larger than the erythrocyte. We must, therefore, not be misguided by the apparently very low concentration of cytochrome *c*, when expressed in the language of the exchange counter. The functional, cellular metabolic concentration is actually appreciable, even judged by the standard of the 'highly concentrated' hemoglobin.

Intracellular position of hemin derivatives. In the mammals, hemoglobin, myoglobin and the cytochrome pigments are essentially intracellular. This again stresses the thought of *non-fixity*. I do not wish to alarm you, but during this symposium (duration 2.5 hours), on the basis of the rate of bile pigment production (18), it may be calculated that each of you will destroy approximately 28,000,000,000 red blood corpuscles. However, it is a relief to know that you will destroy mainly your very old, decrepit cells, those about four months of age (19). During this same interval of 2.5 hours, most of you fortunately, will each produce 28,000,000,000 new, mature, functional erythrocytes. Hence, whatever else you may gain, you should leave this meeting rejuvenated to some extent. On the other hand, this aspect alone of rejuvenation is a metabolic process of large magnitude, involving, in an individual adult per day, the probable conservation and reworking of some 8 gm. of specialized protein (about one-fifth of the total protein requirement), the conservation and reuse of some 0.03 gm. of Fe (at least three times the normal nutritional requirement) and the discarding (elimination as bile pigments) and manufacture *de novo* of about 0.3 gm. of protoporphyrin.

TABLE 1. RELATIONSHIPS OF TOTAL CHROMOPROTEINS TO BODY MASS, W, AND SURFACE AREA, S

SPECIES	CHROMOPROTEIN	CHROMOPROTEIN			
		Total	Per kilo	Per 1 m²	Per $W^{0.75}$
		gm.	*gm.*	*gm.*	*gm.*
Rat (11)	Hemoglobin	3.19	**12.76**	88.3	9.01
W = 0.250 kilos	Myoglobin	0.101	0.404	2.8	0.286
S = 0.0361 m²	Cytochrome *c*[1]	0.0144	0.056	**0.399**	**0.041**
Man (12)	Hemoglobin	900.0	**12.85**	481.3	37.2
W = 70 kilos	Myoglobin	40.0	0.57	21.4	1.65
S = 1.87 m²	Cytochrome *c*[1]	0.780	0.011	**0.417**	**0.032**
Horse (12)	Hemoglobin	5800.0	**11.60**	876.0	55.4
W = 500 kilos	Myoglobin	1868.0	3.74	282.0	17.83
S = 6.62 m²	Cytochrome *c*[2]	16.6	0.033	**2.51**	**0.159**
Horse (12)[3]					
W = 455 kilos	Myoglobin	1347.0	2.96	216.0	13.7
S = 6.23 m²	Cytochrome *c*[2]	24.4	0.054	**3.92**	**0.248**
Cow, heifer (12)	Hemoglobin	2215.0	**12.17**	797.0	44.8
W = 182 kilos	Myoglobin	307.0	1.69	110.3	6.20
S = 2.78 m²	Cytochrome *c*[2]	1.24	0.0068	**0.446**	**0.025**

[1] Values based on complete analyses of individual organs.
[2] Values based on (total muscle cytochrome *c*)/0.8.
[3] Steeplechase thoroughbred, out of training; had been insured for $75,000.

The advantages of cellular confinement become apparent when examined in the light of pathological phenomena such as the *hemoglobinurias* and *met- or ferrihemoglobinemias*. Hemoglobin, myoglobin, and cytochrome *c* were not designed for extracellular residence, and, when these substances beyond certain amounts enter or are placed into the plasma, they can, in spite of their relatively large size (molecular weight), pass into the urine.

In our original experiments (20), hemoglobinuria was studied in the dog, following the intravenous injection of dog hemoglobin. By means of spectrophotometry, the character of the pigment excreted was definitely established, and the blood plasma levels ('threshold') beyond which urinary spillage occurred were determined accurately. The rates of excretion were similar for oxy-, met- and cyanmethemoglobin (20). Oxyhemoglobin was excreted as such, but changed rapidly in the samples obtained by catheterization into a pigment identified as methemoglobin. *Oxidants of hemoglobin are thus present in normal urine.*

This work has been extended (12) to chromoproteinuria in general by a study in the same animal, the dog, of the behavior of intravenously injected horse hemoglobin, horse myoglobin and horse cytochrome *c*, with the respective molecular weights of 66,800, 16,700 and 13,000. The excretion of injected horse hemoglobin and myoglobin was measured also in the horse, while studies are in progress as to the urinary excretion of parenterally administered cytochrome *c* in man. Myoglobinuria in man had become of interest in World War II by its association with the 'crush syndrome' (21), while therapeutic claims had been made (22) for cytochrome *c* in advance of information as to its metabolism. However, a more cogent reason for the investigation was a curiosity aroused by the obvious question of the influence of the size of the molecules upon their penetration of the kidney membrane barriers. A rather subtle and more interesting question was also involved. In the case of hemoglobin no information was available up to its molecular magnitude in the environment of the plasma. Disaggregation—a not wholly theoretical possibility (23)—could have occurred, and it could be argued that the smaller molecules, or only those molecules which had become smaller in the plasma, had leaked through the glomerular filter. The filtration of appreciably larger amounts of myoglobin, a protein very similar in type but only one-fourth the size of hemoglobin, would therefore, be at least presumptive evidence that in hemoglobinuria we are not dealing with an altered molecular species. This very result has been obtained (table 2). As is also evident from the data summarized in this table, the dog, horse and man have approximately the same urinary spillage levels (in plasma) for hemoglobin. Whether these plasma levels may be described as *urinary thresholds* is debatable.

Attention may also be directed to the remarkable urinary excretion of myoglobin in a case of 'Monday Morning Disease', which is not quite what you may think. This is a picturesque and descriptive name of an equine disease, often fatal, which is more formally designated *Paralytic Myoglobinuria*. From the myoglobin excretion in this animal (table 2) and from analyses of the pathological muscle in which a reduction in myoglobin as well as cytochrome *c* content was found, it was calculated that at least 10 per cent of the total muscle mass had been affected. I propose that Monday Morning Disease, while of different etiology, possesses many resemblances to the crush syndrome in man. In the horse, the huge muscles of the hind limbs may be said to be 'crushed' by spastic contracture. Hence, this disease of the horse has more than economic interest for man and is an example of the valuable material available in the veterinary clinic.

Not shown in table 2 are several additional observations of interest. *a*) The concentration of each of the chromoproteins in the first plasma, removed at 5 to 15 minutes after their intravenous injection, was 25 to 45 per cent less than that calculated from the amount injected and the plasma volume (table 2). This early disappearance of the injected chromoprotein from plasma has not been satisfactorily explained, though such factors as blood dilution and rapid distribution over a space larger than the plasma may contribute towards the phenomenon. *b*) Within 2 to 6 hours the chromoproteins were totally removed from the plasma, although (in the case of the injection of hemoglobin in the dog) blood dilution persisted for more than 24 hours. *c*) The amount of urinary chromoprotein was proportional to the amount injected, but, more important, it was an indirect function of the amount which had disappeared from the plasma very soon after intravenous administration (see above). *d*) If the evaluation of chromoprotein 'cleared' from the plasma by way of the urine is based on the total amount in the plasma,

calculated from the concentration in first plasma sample (table 2), rather than from the amount injected, the 'clearance' of hemoglobin is appreciable, while myoglobin 'clearance' is virtually complete. This suggests that myoglobin, which has leaked through the glomerular filter, has not been reabsorbed in tubules, and raises doubts concerning the proposal (24) that hemoglobin (a molecule four times larger) can be reabsorbed by the tubular apparatus, at least by any of the commonly postulated mechanisms. This is one field where isotopically labelled chromoproteins should supply the answer. *e*) In the plasma, hemoglobin was slowly oxidized to methemoglobin, while the injected ferricytochrome *c* was rapidly reduced to ferrocytochrome *c*. The oxidation-reduction potential at pH 7.4 of the plasma is thus physiologically poised between $\sim +0.13$ and $+0.27$ volt, the potentials of the respective chromoproteins, and, in this regard, the extracellular environment is not ideal functionally for either hemoglobin or cytochrome *c* (fig. 1).

Ferrihemoglobin reductive function of erythrocytes. We have been discussing the usefulness of the erythrocytes in keeping the hemoglobin 'in'. This may be regarded as an essential but passive function. A study of the reversion of ferri- or methemoglobin, MHb, to ferrohemoglobin, Hb, and, particularly, of the inhibition of this process (25) have disclosed an active functional mechanism in the so-called 'dead' red blood cells (17). Providentially, the erythrocytes consume very little oxygen for their own needs (17), but they possess the necessary enzymic equipment to maintain hemoglobin in an active reduced state. Agreement has been lacking upon the amount of MHb present normally in blood. Our direct spectrophotometric measurements of the oxygen saturation of the arterial blood of man (26) have led us to conclude that the equilibrium $HbO_2 \rightleftharpoons Hb \rightleftharpoons MHb$ is usually poised at 0.5 per cent or less of MHb (fig. 1). The same value has been found for normal subjects by Van Slyke *et al.* (27). Nonetheless inactive ferrihemoglobin may

TABLE 2. CHROMOPROTEINURIA, FOLLOWING INJECTION OF HEMOGLOBIN, Hb_4, MYOGLOBIN, Hb AND CYTOCHROME *c*, *C*

W = body mass in kilos; *V* = plasma volume in ml.

SPECIES	CHROMOPROTEIN INJECTED	AMOUNT INJECTED	CHROMOPROTEIN			PERCENTAGE OF CHROMOPROTEIN EXCRETED
			First plasma sample	Lowest spillage level (plasma)	Highest concentration in urine	
		mM	*mM/l.*	*mM/l.*	*mM/l.*	%
Dog (20) *W* = 12.8 *V* = 512	Dog Hb_4	0.1365	0.1632	0.0925	0.3785	7.3
Dog (20) *W* = 14.0 *V* = 560	Dog Hb_4	0.3383	0.4110	0.0529	1.175	33.6
Dog (12) *W* = 8.7 *V* = 348	Horse Hb_4[1] " *Hb*[1] " *C*	0.0291 0.0109 0.0344	0.0683 0.0231 0.0560	0.0320 0.0025 0.0014	0.1340 3.890 1.868	4.2 70.3 48.7
Horse (12) *W* = 400 *V* = 23,400 (Monday Morning Disease)[2]	Horse Hb_4[1] " *Hb*[1]	6.160 0.1103	0.1892 0.0029	0.1432 0.0006 0.002?	0.3442 0.1140 1.293[3]	0.3 43.0
Man (12) *W* = 50 *V* = 2250	Human Hb_4[1] Horse and cow *C*	0.419 0.113	0.1328 0.0267	0.0780 0.0009	0.4770 0.8540	9.7 17.2

[1] Pure, salt-free, vacuum dried from the frozen state.

[2] Equine paralytic myoglobinuria or azoturia.

[3] This corresponds to 19.43 gm. myoglobin excreted in 1 day in 900 ml. of urine, an abnormal low volume for the horse. In a three-day period approximately 60 gm. of myoglobin were excreted by this animal.

be found in large amounts abnormally. The results of the *in vitro* study on dog blood of the reduction of intracellular MHb (fig. 2) implicated enzymes concerned with glycolysis, and it was postulated that the DPN and TPN systems could be involved (25). Later Gutman *et al.* (28) furnished direct evidence for the rôle of DPN in this interesting and probably physiologically important process. These findings invite a fresh orientation in regard to concepts of methemoglobin production. Methemoglobinemia need not be an expression (perhaps only rarely is) of intoxication by an oxidant of hemoglobin. It may indicate a damage of the enzymic methemoglobin reductive mechanism, or, indeed, a fundamental defect—a 'biochemical lesion'—as has been suggested recently in cases of idiopathic, familial methemoglobinemia (29).

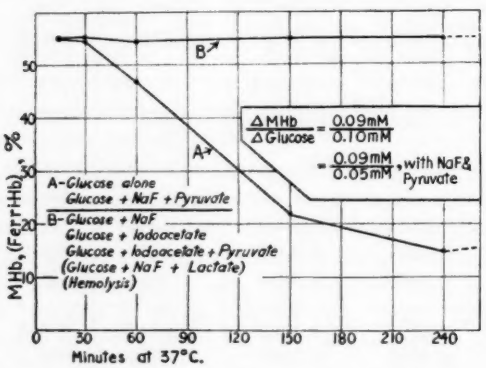


Fig. 2. ENZYMIC REVERSION, and its inhibition of intracellular ferri- or methemoglobin, MHb.

Aspects of the metabolism of cytochrome c. We have used a type of 'depletion technique', partial hepatectomy, in a systematic study of the metabolism of cytochrome *c* in the rat (11, 30–32). This species is particularly suitable for this purpose, since it possesses a discretely multilobed liver, with reproducible ratios of mass of individual lobes to each other (11), and regeneration, following excision of two thirds of the liver, is rapid. In introducing this technique for metabolism studies, it was pointed out (11) that it was a two-edged sword, yielding information not only upon the metabolite under investigation but also upon the restoration process *per se*, in itself of fundamental interest.

Our earlier results permitted the following deductions (30, 31): *a*) The amount of restoration of liver tissue and the quantity of new cytochrome *c* in liver were direct functions of the amount of tissue removed. *b*) Most of the liver regeneration and all of the new cytochrome *c* 'production' occurred in the first 4 to 6 days after liver lobectomy. The rate of appearance of new cytochrome *c* was found to be 16.1, 10.9 and 4.8 per cent per day for restoration periods of 4, 6 and 14 days, respectively. *c*) The appearance of new cytochrome *c* as well as of ribose nucleic acid, PNA, was relatively independent of dietary protein. The concentrations of each were, significantly, appreciably greater in the regenerating livers from rats on no protein than from those on a high 31 per cent protein diet (30). Apparently partial hepatectomy plus protein starvation offered a maximal deprivation stimulus. The experiments furnished evidence, previously not obtained either by quantitative chemical analysis or *in vivo*, in favor of the view that PNA is involved in cellular protein synthesis in tissue restoration (30). *d*) A remarkably great liver regeneration (of the order of 170 per cent) was obtained in animals which were shifted, during restoration, from the no protein to the high protein regimen. In these experiments, the results indicated that most of the cytochrome *c* and PNA had been laid down during the period of protein deprivation. When dietary protein was then made available, tissue regeneration 'went to town'.

These studies have led to the conclusion (30) that 'certain cellular components, like cytochrome *c* and PNA, are preferentially produced or deposited in tissues, and are important or essential in growth and proliferative processes, which appear to depend on *intrinsic* (tissue) as well as *extrinsic* (dietary) factors'.

In a separate group of experiments (31) the effects of parenterally administered cytochrome *c* and of anoxia upon liver regeneration and cytochrome *c* in liver were studied. *a*) Unequivocal evidence was not obtained for the incorporation of injected cytochrome *c* in the regenerating liver, in which a maximal opportunity for such a process should be afforded. On the other hand, in a limited number of experiments increased liver regeneration was found in the cytochrome *c* injected rats. This unexpected finding must be cautiously interpreted, at present, as a nonspecific effect. Anoxia of oxygen deprivation, induced by residence at a simulated high altitude of 15,500 feet (only a moderately high altitude for rats), was without effect either on cytochrome *c* metabolism or on liver regeneration. However, under

these anoxic conditions there was a marked accommodation of the blood oxygen transport system, reflected in large increases of the red blood corpuscles and hemoglobin. Myoglobin was not affected. Thus, although there is doubtless a limit to the effectiveness of this mechanism of adjustment, in these experiments there was a clear separation of a *primary response* to oxygen deficit in the inspired air from any postulated *secondary response*, involving the oxygen utilization mechanism, the cytochrome system, in the tissues (fig. 1). If hemoglobin and the circulation respond adequately to oxygen deficit, there is no need for secondary adjustment in oxygen utilization. Indeed, from a teleogical standpoint, the body would lack physiological wisdom if in the face of a diminished oxygen supply, it responded by a greater use of oxygen in the tissues. A temporary advantage might be gained from such a process, but it is a spendthrift's act of desperation, a luxurious meal, which completely empties the diminished stores in the cupboard. Support for this view has been furnished in our finding that deeply anesthetized dogs (at low levels of metabolism and, hence, of oxygen requirement) were able to withstand, without any damage to the central nervous system, blood levels of 93 per cent carbonyl hemoglobin (33), whereas unanaesthetized animals had 'critical' levels of about 75 per cent HbCO (34).

Hormonal influence on cytochrome c. The cytochrome *c*-body mass relationship, which has been discussed, and the greater concentration of cytochrome *c* in tissues with higher rates of oxygen consumption (35) directed our attention to an investigation of the influence of hormonal factors on cytochrome *c*. The literature contained a scattering of inconclusive information (36, 37), some of it questionable (37) owing to limitation of results to a single tissue (muscle) and unacceptably low analytical values for cytochrome *c* concentration in normal tissue. To be of significance, a hormonal effect upon cytochrome *c*, a constituent of all aerobic cells, should be demonstrated for all tissues. Table 3 is a summary of mean values obtained by us (32) for the concen-

TABLE 3. EFFECT OF THYROIDECTOMY, THIOURACIL AND THYROXINE ON CYTOCHROME *c* (32)
Rats of 200 to 250 gm. body weight on high protein diet (30); thiouracil, 50 mg. per day; thyroxine subcutaneously, 1 mg. every other day.

EXPERIMENT	CYTOCHROME *c*				CYTOCHROME *c* IN RESTORED LIVER		LIVER RESTORATION	LIVER PNA	LIVER DNA
	Liver	Kidney	Heart	Muscle	Total	New pigment			
	γ/gm.[1]	γ/gm.	γ/gm.	γ/gm.	γ	%	%	mg./gm.	mg./gm.
Thyroidectomized									
35 days before liver lobectomy	138 ±5[3]							7.19 ±0.10	2.80 ±0.05
14 days after liver lobectomy	181 ±5	248 ±12	316 ±16	57 ±3	780	68.9	65.0	8.35 ±0.11	2.79 ±0.06
Thiouracilized									
45 days before liver lobectomy	145 ±3							8.35 ±0.21	2.70 ±0.03
14 days after liver lobectomy	165 ±3	250 ±5	331 ±4	55 ±3	776	66.1	80.2	9.82 ±0.40	2.74 ±0.05
Thyroxinized, for 14 to 22 days	247 ±6	422 ±8	618 ±11	133 ±6				11.08 ±0.13	2.60 ±0.08
Controls									
	178 ±4							8.50 ±0.25	2.46 ±0.02
14 days after liver lobectomy	210 ±5	352 ±21	447 ±16	98 ±6	1325	67.6	73.9	6.96 ±0.24	3.53 ±0.04

[1] Wet weight of tissue. [2] 68.4 per cent of liver excised. [3] Values after ± are standard errors.

tration of cytochrome *c* in liver, kidney, heart and skeletal muscle in groups of rats subjected to thyroidectomy, thiouracil thyrotoxicosis (withdrawal of thyroxine by interference with its production) and acute hyperthyroidism induced by injection of thyroxine, in comparison with normal controls. The data appear consistently and conclusively to support the thesis of a relationship of thyroid function to cytochrome *c* in tissues. After both thyroidectomy and thiouracil intoxication there was a striking reduction in total body cytochrome *c*, which was reflected (table 3) finding of more pronounced liver restoration after thiouracil. This effect was obtained in all the individuals in this group. It may be noted that in the thyroidectomized, thiouracilized and thyroxinized rats the changes in liver PNA are in the same direction as the changes in cytochrome *c*.

It is tempting to draw upon these observations for a concept of how thyroxine regulates the rate of oxygen consumption, since this phenomenon has remained unexplained. The unfolding of the full story must await developments in a study of the influence on cytochrome *c* of other hor-

TABLE 4. LOCALIZATION OF CYTOCHROME *c* IN LIVER CELL (12)

FRACTION	W:D		CYTOCHROME *c*			DISTRIBUTION OF CYTOCHROME *c*	DISTRIBUTION OF CELLULAR MASS	
	a[1]	*b*[2]	per gm. wet weight	per gm. dry weight	Total			
			γ	γ	γ	%	gm.[3]	%
Liver suspension	18.45	21.10	21.3	393	4155		8.53	
Unbroken cells (600 × gravity)	7.53	7.78	70.0	527	1895		3.48	
				difference	2260	100.	5.05	100.
Nuclear phase (1500 × gravity)	11.90	12.75	0.	0	0	0.	0.51	10.1
Mitochondrial phase (9400 × gravity)	5.58	5.67	263.	1470	1605	71.0	1.07	21.2
Supernatant, microsomal phase[5]	36.1	49.10	3.42	123	494	21.8	2.92	57.9

[1] *a* = wet weight to dry weight of original material.

[2] *b* = wet weight to dry weight, corrected for added NaCl.

[3] Dry weight of tissue.

[4] 204 ml. of 0.85% NaCl solution containing 34.76 gm. of liver parenchyma, recovered from 55.75 gm. of whole liver, with an original cytochrome *c* content of 9446 γ. The analytical values are for 180 ml. of the liver suspension, containing 30.62 gm. of parenchyma. The fibrous portion of liver was 20.99 gm. with 3550 γ of cytochrome *c*.

[5] Washings of mitochondrial phase added to supernatant.

in statistically significant decreases in cytochrome *c* concentration (as well as content) in all the tissues examined, although the changes were of greater magnitude in skeletal muscle than in other tissues. The administration of thyroxine had the opposite effect, significant increases in cytochrome *c*.

In the thyroidectomized rats, liver regeneration was only slightly less than normal, and the increase in cytochrome *c* concentration, characteristic of regenerating liver (11, 30, 31), was also exhibited in the thyroidectomized and thiouracilized animals (table 3). An unusual result is the mones, i.e., of the adrenal cortex and possibly of the pituitary. However, our findings permit a tentative postulate: Thyroxine exerts its effect through the agency of cytochrome *c*. As a working hypothesis, this offers a first approximation in a rational chemical explanation of the mechanism of hormonal control of cellular oxygen utilization, in which thyroxine must have a major rôle. It has the virtue of simplicity in defining the locus of the action of this hormone.

The cellular localization of cytochrome c. Several observations invited inquiry as to the state of cytochrome *c* in cells. 1. Myoglobin is readily

extracted with water from ground muscle tissue, which does not release cytochrome *c* to this solvent. The latter pigment remains in the 'press cake', from which (or from the original tissue) it may be isolated, but trichloroacetic (38) or sulfuric acid (14) must be used as extracting agents. These simple facts are interesting. They suggest that myoglobin may be relatively 'free', whereas cytochrome *c* may be 'bound' in the cellular structure.

2. The activity (in terms of oxygen uptake) of cytochrome *c* in intact tissues (35) is of an appreciably higher order of magnitude than that obtained by the addition of the best cytochrome *c* preparations to tissue homogenates (39)—a fact insufficiently recognized.

3. In the regenerating liver, as has been pointed out, parallel changes in cytochrome *c* and PNA were found (30). By means of the differential centrifugation technique developed in Claude's laboratory, PNA had been shown (40) to be a constituent of cytoplasm, and cytochrome oxidase was demonstrated to be localized exclusively in the cytoplasmic large particle aggregates, the mitochondria (41). This suggested the possibility that cytochrome *c* might be similarly localized (30). The idea is now beyond the speculation stage. While we were engaged in testing it by applying Claude's methods, Schneider, Claude and Hogeboom (42) proved the point in their very recent contribution.

Our tally sheet for the internal cellular distribution of cytochrome *c* is supplied in table 4. The results essentially confirm those of Schneider *et al.*, although we find an appreciably higher localization in the mitochondrial phase, 71 per cent in the large aggregate fraction after washing with 0.85 per cent NaCl as against their 40 per cent (42). It also appears probable that at least some (if not most) of the microsomal phase may be an artifact, representing the breakdown of the large aggregates. If that is the case, nearly all of the cytochrome *c*, like cytochrome oxidase, may have been *in vivo* localized in the mitochondria. This would be consonant with our finding (35) that the determination of cytochrome *c* can be used as an index of the activity of cytochrome oxidase in tissues. Aside from other considerations, such a localization increases manyfold the effective reacting concentrations of cytochrome *c*, the substrate, and cytochrome oxidase, the enzyme (fig. 1).

Derived from an application of the newer cytochemical techniques, particularly that of Claude, is a fascinating picture of the mitochondria, floating batteries carrying cargoes of enzymic charges on a sea of cytoplasm. Enticing vistas are opened. The avenues of investigation of cytochrome *c* converge—its relationship to body mass, the proportionality of its cellular concentration to the oxygen consumption of the tissue, its intracellular orientation with its oxidase and the suggestive nature of its hormonal control.

A large part of our work on the metabolism of cytochrome *c* has been done under contract between the Office of Naval Research and the University of Pennsylvania.

REFERENCES

1. Barcroft, J. *The Respiratory Function of the Blood.* Part II. Hemoglobin. Cambridge, England, 1928. Van Slyke, D. D. and J. M. Neill. *J. Biol. Chem.* 61: 523, 1924. Bock, A. V., H. Field, Jr. and G. S. Adair. *J. Biol. Chem.* 59: 353, 1924.
2. Taylor, J. F. and A. B. Hastings. *J. Biol. Chem.* 131: 649, 1939.
3. Drabkin, D. L. In O. Glasser, Photometry and Spectrophotometry *Medical Physics*, Chicago, 967, 1944.
4. Coryell, C. D., F. Stitt and L. Pauling. *J. Am. Chem. Soc.* 59: 633, 1937. Coryell, C. D. and L. Pauling. *J. Biol. Chem.* 132: 769, 1940.
5. Svedberg, T. and K. O. Pedersen. *The Ultracentrifuge.* Oxford, 1940.
6. Drabkin, D. L. *J. Biol. Chem.* 164: 703, 1946.
7. Perutz, M. F. *Nature* 143: 731, 1939; 149: 491, 1942; 150: 324, 1942. Boyes-Watson, J. and M. F. Perutz. *Nature* 151: 714, 1943. Perutz, M. F. and O. Weisz. *Nature* 160: 786, 1947.
8. Fischer, H. and H. Orth. *Die Chemie des Pyrrols.* Vol. II. Leipsig, 1937.
9. Drabkin, D. L. *J. Biol. Chem.* 158: 721, 1945.
10. Vickery, H. B. and A. White. *Proc. Soc. Exp. Biol. Med.* 31: 6, 1933. Vickery, H. B. *J. Biol. Chem.* 144: 719, 1942. Theorell, H. and Å. Åkesson. *J. Am. Chem. Soc.* 63: 1804, 1941.
11. Crandall, M. W. and D. L. Drabkin. *J. Biol. Chem.* 166: 653, 1946.
12. Drabkin, D. L. Unpublished observations.
13. Millikan, G. A. *Physiol. Rev.* 19: 503, 1939.
14. Rosenthal, O. and D. L. Drabkin. *J. Biol. Chem.* 149: 437, 1943.
15. Kleiber, M. *Physiol. Rev.* 27: 511, 1947.
16. Drabkin, D. L. *Federation Proc.* 4: 88, 1945; *Am. J. Med. Sci.* 209: 268, 1945; 215: 110, 1948.
17. Drabkin, D. L. *Science* 101: 445, 1945.

18. Sribhishaj, K., W. B. Hawkins and G. H. Whipple. *Am. J. Physiol.* 96: 449, 1931.
19. Shemin, D. and D. Rittenberg. *J. Biol. Chem.* 166: 621, 627, 1946. Jope, E. M. *British J. Ind. Med.* 3: 136, 1946.
20. Drabkin, D. L. *Proceedings of the Fifth Summer Conference on Spectroscopy and Its Applications, 1937.* New York and London, 94, 1938.
21. Bywaters, E. G. L., G. E. Delroy, C. Rimington and J. Smiles. *Biochem. J.* 35: 1164, 1941. Bywaters, E. G. L. and J. K. Stead. *Quart. J. Exp. Physiol. and Cognate Med. Sci.* 33: 53, 1944.
22. Proger, S., D. Decaneus and G. Schmidt. *J. Clin. Invest.* 24: 864, 1945.
23. McFarlane, A. S. *Biochem. J.* 29: 660, 1935.
24. Oliver, J. *Harvey Lect.* 40: 102, 1944–45.
25. Drabkin, D. L. *Federation Proc.* 5: 132, 1946.
26. Drabkin, D. L. and C. F. Schmidt. *J. Biol. Chem.* 157: 69, 1945.
27. Van Slyke, D. D., A. Hiller, J. R. Weisiger and W. O. Cruz. *J. Biol. Chem.* 166: 121, 1946.
28. Gutman, H. R., B. J. Jandorf and O. Bodansky. *J. Biol. Chem.* 169: 145, 1947.
29. Barcroft, H., Q. H. Gibson, D. C. Harrison, and J. McMurray. *Clin. Sci.* 5: 145, 1945. Graybiel, A., J. L. Lilienthal, Jr. and R. L. Riley. *Bull. Johns Hopkins Hosp.* 76: 155, 1945. Finch, C. A., H. A. Eder and R. W. McKee. *Proc. Am. Soc. Clin. Invest., J. Clin. Invest.* 25: 293, 1946.
30. Drabkin, D. L. *J. Biol. Chem.* 171: 395, 1947.
31. Drabkin, D. L. *J. Biol. Chem.* 171: 409, 1947.
32. Drabkin, D. L. *Federation Proc.* 7: 151, 1948.
33. Rosenthal, O., H. Shenkin and D. L. Drabkin. *Am. J. Physiol.* 144: 334, 1945.
34. Drabkin, D. L., F. H. Lewey, S. Bellet and W. H. Ehrich. *Am J. Med. Sci.* 205: 755, 1943. Lewey, F. H. and D. L. Drabkin. *Am. J. Med. Sci.* 208: 502, 1944.
35. Rosenthal, O. and D. L. Drabkin. *Federation Proc.* 2: 69, 1943; *J. Biol. Chem.* 150: 131, 1943.
36. Tipton, S. R. *Endocrinology* 34: 181, 1944. Tipton, S. R., M. J. Leath, I. H. Tipton and W. L. Nixon. *Am. J. Physiol.* 145: 693, 1945–46.
37. Tissières, A. *Arch. intern. physiol.* 54: 305, 1946.
38. Keilin, D. and E. F. Hartree. *Proc. Roy Soc.* (London) B, 122: 298, 1937.
39. Elliott, K. A. C. and M. E. Greig. *Biochem. J.* 32: 1407, 1938.
40. Schneider, W. C. *J. Biol. Chem.* 165: 585, 1946.
41. Hogeboom, G. H., A. Claude and R. D. Hotchkiss. *J. Biol. Chem.* 165: 615, 1946.
42. Schneider, W. C., A. Claude and G. H. Hogeboom. *J. Biol. Chem.* 172: 451, 1948.

METABOLISM OF IRON

PAUL F. HAHN

From Meharry Medical College, Nashville, Tennessee

Our knowledge of the metabolism of iron has been greatly increased during the past decade largely due to the advent of the radioactive isotopes of this metal. Iron is present in the body in a wide variety of forms of combination. By purely chemical methods alone it had proved practically impossible to trace the course taken by recently ingested iron. The presence of larger amounts of it in various forms in the tissues under examination masked the presence of the relatively minute amount added as a result of feeding. As in the case of other metabolites, tracer studies with isotopes led to the solution of many of these problems. Iron plays a vital part in the metabolic activity of every mammalian body cell. Cytochrome, catalase and muscle hemoglobin iron are husbanded by the body regardless of the stresses imposed by serious deficiency disease. Some other iron fractions such as plasma iron, on the other hand, are extremely labile. The bulk of the body iron exists in the red blood cell hemoglobin and much information has been and promises to be learned about the formation and breakdown of this compound. Application of the newer methods of research in this field requires a high degree of collaborative effort on the parts of the physicist, chemist, biologist and clinician. Such collaboration probably represents the pattern of much future medical research.

In no sense shall I attempt to review here what has been done in the whole field of iron metabolism. Rather I shall attempt to bring out the significance of a few of the more recent developments and to point out a few of the more glaring deficiencies in our present state of knowledge.

One of the first contributions made with the use of the Fe^{59} isotope was related to the unique ability of the body to accept or reject iron following ingestion, depending on the need for the element (11). Selective absorption had been suspected as a result of the findings of Welch, Wakefield and Adams (33) in which no excretion of iron was demonstrated in a patient with an ileostomy stoma and an isolated colon. McCance and Widdowson (23) later showed that even following the injection of considerable amounts of iron by vein, there was a negligible amount of excretion. On the basis of these balance studies they had the courage to suggest that selective absorption prevented an accumulation of undesirable amounts of this element in the body. With the isotope it was possible to recognize the presence of recently administered iron in plasma and later in the red cell hemoglobin and to show that in iron-deficiency anemia the absorption was increased many times over the normal absorption. This finding has been corroborated many times since in our and other laboratories. That anemia *per se* did not influence the uptake of the metal was seen from the fact that normal dogs bled acutely about 60 per cent of their estimated blood volumes when fed tagged iron absorbed no more than when in the normal state. However, if a week or so were allowed to elapse following the hemorrhage, the uptake of tagged iron was several times what it had been (10, 17). This finding led us to postulate an acceptor mechanism in the gastro-intestinal mucosa which normally was saturated with iron and, therefore, allowed a minimal of passage from the lumen to the plasma. The interesting work which was currently being done by Granick and Michaelis (6, 7, 24) on the properties of ferritin pointed to the possibility of that compound being the intermediary responsible for this unusual phenomenon. Accordingly, a joint study between the Rochester and Rockefeller Institute laboratories was set up and it was shown that inorganic iron could be converted readily to ferritin iron in the liver. Further, it was demonstrated that hemoglobin iron upon the destruction of red cells by acetylphenylhydrazine was in part, at least, converted to ferritin in the liver and spleen. Thus the storage function of ferritin was established (14). Granick continued the work (8, 9) and was able to demonstrate larger amounts of ferritin in the duodenum than in any other part of the G-I tract. Furthermore, he discovered that there was an increase in the amount of ferritin in the intestinal tract when animals were fed iron repeatedly. This very interesting finding suggests an ability of the body to acclimate itself to iron feeding and deserves

considerable study since such a mechanism might be an important factor in the efficiency of therapeutic measures. Unfortunately, little has been done to date in studying the comparative uptakes of single or multiple doses of iron. Such studies could only be carried out successfully with the isotope method.

Distribution of iron. No studies of the true distribution of iron in the human body have been carried out up to this time. Chemical determinations have been obscured by the presence of blood in tissues. However, using the viveperfusion technique values have been obtained for dogs by Bogniard and Whipple (2) and Hahn and Whipple (19) and for rats by Austoni, Rabinowitch and Greenberg (1). For the distribution in dogs see table 1.

TABLE 1

	% TOTAL BODY IRON
Blood hemoglobin iron	57
Muscle hemoglobin iron	7
Total hemoglobin iron	64
Parenchyma iron (catalase, cytochrome, peroxidase, etc.)	16
Available visceral storage (ferritin, hemosiderin, etc., in liver, spleen, and bone marrow)	15
Available iron, other tissues (estimated)	5
Total iron	100

These values were arrived at by analysis of dogs of about 20 kg. weight with a circulating blood volume of about 1500 ml. and a bulk of striated muscle estimated at 6.5 kg. Since the muscle hemoglobin and parenchyma iron are unaffected by hemorrhage or dietary depletion, the available stores of iron are such as to allow a 30 per cent replacement of the blood hemoglobin in time of need.

Efficiency of iron absorption. Fed at anything other than ridiculously low levels the absorption of iron is notoriously inefficient. When chronically anemic human subjects or animals are fed doses of tagged iron at 0.2 to 1.0 mg. levels there may be as much as 50 to 90 per cent uptake (18). When levels such as might occur in the normal dietary are fed, i.e., 5.0 mg., the growing child was found to absorb only about 12 per cent (4). In the latter stages of pregnancy when the need is augmented, 5.0 to 18.0 mg. doses of the tagged iron administered result in the uptake of only about 35 per cent (13).

Too little attention has been paid by clinicians to the work of Whipple and Robscheit-Robbins (34) in this matter. Many years ago they showed that in their standardized anemic dogs iron fed at a daily level of 40 mg. over a two-week period was absorbed and utilized to the extent of 35 per cent. When the dosage level was increased to 400 mg. per day the percentage uptake was only 5 to 6 per cent. Thus, an increase in dosage level by a factor of ten resulted in less than double the absolute uptake. These findings are in agreement with those obtained using single feedings of tagged iron. Since the commonly used therapeutic dose is 325 mg. one may expect only about 16 mg. of actual iron absorption. This is the equivalent of about 5 gm. of hemoglobin. The normal adult male has about 750 gm. of blood hemoglobin in circulation and this may frequently be reduced to one third or less by acuteor chronic hemorrhage. Thus, it becomes apparent that iron must be taken assiduously for many weeks in order to make up losses due to massive acute hemorrhage.

At present there is nothing known to enhance the absorption of iron. On the other hand, many reports indicate that phosphates, phytates and other dietary ingredients reduce the absorption. It would, therefore, seem advisable to administer iron between meals when given either prophylactically or therapeutically.

Valence state and iron uptake. Clinical experience had shown that iron in the ferrous form seemed to be a more suitable means of providing the patient with iron than the corresponding ferric salts. That the latter salts are poorly tolerated was well known and probably influenced the physician considerably in his thinking and practice. However, this presumption was probably largely based on empiricism. Whipple and Robscheit-Robbins (34) had found in their multiple feeding experiments in dogs that there was little if any difference in uptake whether ferrous, ferric or reduced iron was administered. In single feeding experiments in both dogs and human subjects we found by using the tagged iron method that ferrous salts were many times more efficiently absorbed (15, 16). Moore and his colleagues (27), however, found ferrous iron more readily absorbed in human patients but noted no difference in the uptake in dogs. The latter group have suggested that a species difference may exist,

but we feel that our data shows conclusively that the disparity obtains in both species.

From a more practical aspect it has been possible to demonstrate that the need for iron determines its uptake. Feeding a standard test dose of tagged iron to 176 school children in the 7 to 10-year age group, we found that the uptake was closely correlated (4) with Heath and Patek's (20) estimated yearly increases in body iron during growth.

More recently in a Cooperative Study of Infant and Maternal Nutrition conducted in Nashville during the past two years, nearly 1000 pregnant women were fed tagged iron. It was found that there was from 50 to 100 per cent greater absorption of iron in the last quarter of pregnancy than in the first quarter (13). This is an excellent reflection of the fetal growth demand. Further, it indicates that the logical time for prophylactic and replacement iron medication is in the last half of pregnancy, at which time the iron will be most efficiently taken up. Prescribing iron early in pregnancy should reasonably be avoided since the saving from lack of catemenic blood loss will more than repay the early fetal needs. Furthermore, a woman's good intentions to take her thrice daily pills will have long since been forgotten when the supplement would do the most good.

TABLE 2. UPTAKE OF RADIOACTIVE IRON AT VARIOUS STAGES OF GESTATION IN 329 PREGNANT WOMEN

NO. PTS. IN GROUP	UPTAKE AT WEEKS OF GESTATIONAL PERIOD			
	0–10	11–20	21–30	31–40
22	18.2			
122		21.4		
108			31.5	
77				38.5
329				

Transport of iron. That the blood plasma is the vehicle of iron transport is now generally agreed. On being absorbed from the gastro-intestinal tract it is felt that the iron is oxidized to the ferric form and combined firmly with plasma globulin (21, 31). This transport function has been studied extensively by Moore and his group (25, 26), Waldenström (32), Laurell (22) and by the use of the radioactive isotope by Hahn, Bale, Lawrence and Whipple (11) and Yoshikawa, Hahn and Bale (35). Depletion of the labile stores has been shown to be reflected in a lowered serum iron level (21, 22, 26). Normal values may range from 80 to 200 μg. per 100 ml. In iron deficiency anemia or pernicious anemia in remission, the serum iron may be markedly lowered, whereas in pernicious anemia in relapse, it may be somewhat elevated.

Many attempts have been made to quantitate the amount of iron absorption from the intestinal tract by means of the rise in plasma iron levels. However, it should be kept in mind that at least two processes are occurring simultaneously, namely, passage from the lumen of the gut to the bloodstream and removal from the latter by the storage depots and the hemopoietic system. That these processes may have different rates under various circumstances seems altogether likely and, therefore, such quantitation attempts are probably highly unreliable.

At this writing, no clear-cut demonstration of the transport of iron from the storage depots to the bone-marrow has been made.

The importance of the plasma iron fraction should not be underestimated merely because of the relatively small amounts appearing in the circulation at any one time. The alterations in level mentioned above are of great clinical significance. The depression of plasma iron levels in inflammatory and infectious processes is under study by Cartwright, Wintrobe and their collaborators (3) using the radioactive isotope and should add considerably to our knowledge of the metabolism of this element in disease.

Inflammation and iron absorption and utilization. In cachexia, many inflammatory disorders and infectious disease, there is a high incidence of accompanying anemia. This condition, though generally recognized, is not at all understood. Some think that toxic factors are at play with a resultant depression of the bone-marrow activity. It is well known on an empirical basis that iron therapy is of no value while the disease process is rampant.

Robscheit-Robbins and Whipple demonstrated in a dog with chronic endometritis a markedly decreased rate of utilization of iron (30). Following hysterectomy there was a resumption of the usual hemoglobin production.

Using tagged iron and the turpentine sterile abscess as a means of inducing a controllable inflammatory process, we were able to show that an abscess induced in the axillary region was able

to depress iron absorption considerably (12). The results of this set of experiments can be seen in table 3. Some might argue that there had been absorption without subsequent utilization of the iron. Therefore, in the case of one animal, the actual amounts of total radioactivity removed by repeated bleeding was determined and was found to agree very favorably with the difference in circulating amounts of tagged iron before and after bleeding. This indicated that there had been no storage of the fed isotopic iron.

TABLE 3. IRON ABSORPTION BY DOGS WITH SUBCUTANEOUS TURPENTINE ABSCESSES

DOG NO.	HEMATOCRIT	FORM ADMINISTERED	DOSE	UPTAKE	EXPERIMENT TYPE
	%		*mg.*	%	
40–149	13.4	$FeCl_2$	9	41	Control
"	23.0	"	10	4.1	Turpentine 1 day previously
"	18.3	"	10	18.1	Control
41–164	26.9	Fe Am Cit	38	4.1	Turpentine 2 days previously
"	28.6	"	34	17.6	Control
37–196	22.4	"	38	5.3	Turpentine 2 days previously
"	23.4	"	34	6.9[1]	Control

[1] This animal had been splenectomised and, therefore, was subject to Bartonella infection. That this infection was responsible for lowered uptake at this time as suspected by clinical condition seemed indicated by reaction to treatment with Mapharsen.

Balance sheet for first feeding of *dog 37-196*
10 days after feeding, Hct. 35.2%;
Estimated total circulation radio iron
4720 cpm.
2 weeks later, Hct. 28.7%;
Estimated total circulation radio iron
635 cpm.

Radioactivity lost to circulation. . . . 4085 cpm.
Radioactivity found in 1900 ml. blood removed in two-week interval. . . . . 4045 cpm.
This balance shows that with the turpentine abscess there was no additional absorbed and stored iron available for use under the stimulation due to blood loss.

Recently Gibson and Finch (5) have been studying the effect of infection on iron utilization of tagged iron administered intravenously in human subjects. "The results obtained showed that the utilization of iron for hemoglobin synthesis was impaired to a degree roughly proportionate to the severity of the infection as judged by clinical evidence."

Thus we see that inflammation and infection interfere both with the absorption and utilization of iron. Much remains to be done in elucidating the mechanisms of these reactions.

Pathway of iron from red cell hemoglobin breakdown. It seems reasonable to allow a little speculation to enter the scene at this point regarding the path of breakdown of hemoglobin. From the work of Lemberg, Whipple and Barkan and their associates, it would be inferred that the picture would be somewhat as follows: Upon completion of its four-month life cycle the red cell having deteriorated in some unknown manner would be phagocytosed by the cells of the lymphoid-macrophage system. The released hemoglobin would then undergo splitting at the alpha methene linkages of the porphyrin part of the molecule giving rise to verdo-hemoglobin. The iron would then be split off and stored in the Kupfer cells and macrophages for future use. The globin would then become dissociated from the porphyrin derivative and enter the protein metabolic pool. The bilirubin would be quantitatively secreted by the biliary tract. Another possible pathway presents itself, suggested by some rather provocative data accumulated over the years in studying the iron isotope. Under conditions of red cell destruction due to hydrazine derivatives it appeared that iron from recently destroyed red cells was utilized preferentially, regardless of the amount of readily available storage iron that might be present in the liver, spleen and bone marrow. Also, as has been pointed out, it should be kept in mind that under these conditions tagged red cell iron has been shown to be deposited in the form of ferritin. This latter material is an iron-protein compound containing up to 23 per cent of its weight of the metal and having a molecular weight of 460,000 (9). Under conditions of acute or chronic blood destruction pathologists have for many years recognized iron in the form of hemosiderin by the Prussian blue staining reaction. Since hemosiderin and ferritin are laid down under the same conditions in the same cells, it is only reasonable to inspect the

possibility of their being either identical or closely related. Simplification of our view of the path taken during the destruction of hemoglobin could be accomplished if it were not for a few observations. Globin has a measured molecular weight of about 68,000. This does not preclude the possibility, however, that apoferritin might be an aggregate of globin. Much confusion stemmed from the arbitrary classification of proteins through the salting-out technique which is being clarified by modern electrophoretic studies. In this connection it should be well to recall the similarity between serum albumen and globin as shown by Pederson and Waldenström (28).

Tentatively, and at the risk of severe criticism from those much better versed in the chemistry and structure of proteins, I should like to suggest a slightly different pathway of hemoglobin disintegration than that outlined above. Upon phagocytosis of the red blood cell let us suppose that following splitting at the methene linkages the porphyrin entity of the molecule is immediately divorced from the iron-globin residue, becoming attached to serum albumen or not, as the case may be, and undergoing secretion by the hepatic-biliary system. The iron-globin complex would then be deposited as a precursor of hemosiderin and/or ferritin under some conditions, or would be transported to the bone marrow for immediate incorporation into the megaloblast, reticulocyte or normoblast, depending on which of these forms of the red cell are actually the site of hemoglobin formation. Such an hypothesis should lend itself readily to test using N^{15} as a tag in the globin fraction of the hemoglobin molecule.

Problem of iron deficiency anemia. Etiologically, iron deficiency anemia may be classified as follows: *a*) hemorrhagic, acute or chronic; *b*) nutritional; *c*) parasitic, with or without nutritional complications; *d*) gestational, with or without hemorrhagic or nutritional complications; *e*) other obscure or unusual causes.

a) It has been stated repeatedly elsewhere that microcytic, hypochromic anemia does not occur in the normal male adult unless there has been blood loss. The common causes of acute hemorrhage include gastric and duodenal ulcers, ruptured esophageal varices, uterine fibromata, tumors of the gastro-intestinal tract, premature separation of the placenta, etc. These conditions nearly all also contribute largely to the causes of chronic hemorrhage, to which should be added hemorroids, menorrhagia, metrorrhagia, hemorrhagic diathesis due to leukemia, etc. There is still some disagreement as to the advisability of transfusion in acute massive hemorrhage as an immediate measure for fear of inducing further hemorrhage. Each case presents an individual problem. However, there is no doubt that iron therapy is indicated for both massive acute and chronic hemorrhage. Due to the inefficient absorption of iron in oral therapy the element in suitable form must be administered *frequently* and *continuously for many weeks* in order to care for replacement and storage needs. There being no suitable preparation of iron for parenteral use at the present the only means by which large quantities of iron may be administered in a short period is by infusion of whole blood or suspended red cells. When there is a chronic low-grade blood loss it is a simple matter to maintain the patient in equilibrium as long as replacement therapy is faithfully adhered to. For such treatment it is recommended that small doses of 30 to 60 mg. of iron as the ferrous salt be administered at frequent intervals, preferably between meals, over prolonged periods of many weeks.

b) Nutritional iron deficiency anemia is a relatively rare condition in the United States. It is probable that only 5 to 10 mg. of absorbable iron are needed to take care of the growth requirements of children and the catamenic losses in women. The requirement for men is probably only a fraction of this. Since our average dietary contains well over this amount, we see very little uncomplicated anemia due to lack of intake of the element. Where it exists it may be a result of local peculiarities of diet or in individual instances, a result of food fadism. Nutritional anemia in northern Brazil, India and China constitutes a serious problem on the other hand, especially when the deficient dietary is complicated by pregnancy or parasitism.

c) Parasitism may play an important role in iron deficiency anemia as perhaps best exemplified by the indigent population of Brazil. In the northern part of the latter country where the worm burden due to ankylostomiasis may reach 400 or more per person, the incidence of hookworm anemia is very high. Here the dietary consists largely of rice and black beans. On the other hand, although the worm burden in the same class of population in the southern part of Brazil may be equally as high, the incidence of anemia is low. Presumably, this is because of the high beef content of the dietary supplying sufficient replacement of red blood cell hemoglobin iron.

d) Gestational anemia may be due to multiple frequently-occurring pregnancies, especially where there has been no prophylactic iron taken late in pregnancy or where the dietary is inadequate to take care of the extra demands. Other common

causes are excessive hemorrhage at parturition or post-partum.

e) Chronic infection or some other process which interfered with normal iron uptake may, in the presence of unusual demands for iron such as occur during growth or pregnancy, result in an iron deficiency condition. We have seen an instance where pernicious anemia in relapse in a growing child (a condition marked by lowered iron absorption) resulted in a superimposed iron deficiency anemia which became apparent on institution of specific therapy with liver extract.

The replacement requirements for iron therapy can be defined simply. The chief difficulty is in estimating the true blood loss or deficiency. When the latter can be estimated one may apply the value of 0.5 mg. of iron needed to be absorbed for each ml. of blood of normal hematocrit value lost.

REFERENCES

1. Austoni, M. E., A. Rabinowitch and D. M. Greenberg. *J. Biol. Chem.* 134: 17, 1940.
2. Bogniard, R. P. and G. H. Whipple. *J. Exp. Med.* 55: 653, 1932.
3. Cartwright, G. *Blood* 2: 256, 1947.
4. Darby, W. J., P. F. Hahn, M. M. Kaser, R. C. Steinkamp, P. M. Densen and M. B. Cook. *J. Nutrition* 33: 107, 1947.
5. Gibson, J. G. II and C. Finch. Personal communication.
6. Granick, S. *J. Biol. Chem.* 146: 451, 1942.
7. Granick, S. *J. Biol. Chem.* 149: 157, 1943.
8. Granick, S. *J. Biol. Chem.* 148: 463, 1943.
9. Granick, S. *Chem. Rev.* 38: 379, 1946.
10. Hahn, P. F. *Advances in Medical Physics.* New York: Academic Press, 1948.
11. Hahn, P. F., W. F. Bale, E. O. Lawrence and G. H. Whipple. *J. Exp. Med.* 69: 739, 1939.
12. Hahn, P. F., W. F. Bale and G. H. Whipple. *Proc. Soc. Exp. Biol. Med.* 61: 405, 1946.
13. Hahn, P. F., E. L. Carothers, R. O. Cannon, C. W. Shappard, W. J. Darby, M. M. Kaser, G. S. McClellan and P. M. Densen. *Federation Proc.* 6 (1): 1947.
14. Hahn, P. F., S. Granick, W. F. Bale and L. Michaelis. *J. Biol. Chem.* 150: 407, 1943,
15. Hahn, P. F., E. Jones, R. C. Lowe, G. R. Meneely and W. Peacock. *Am. J. Physiol.* 143: 191, 1945.
16. Hahn, P. F., R. C. Lowe, G. R. Meneely and W. Peacock. *Federation Proc.* 3: 89, 1944.
17. Hahn, P. F., J. F. Ross, W. F. Bale, W. M. Balfour and W.M. Whipple. *J. Exp. Med.* 75: 221, 1942.
18. Hahn, P. F., J. F. Ross, W. F. Bale and G. H. Whipple. *J. Exp. Med.* 71: 731, 1940.
19. Hahn, P. F. and G. H. Whipple. *Am. J. Med. Sci.* 191: 24, 1936.
20. Heath, C. W. and A. J. Patek. *Medicine* 16: 267, 1937.
21. Heilmeyer, L. and K. Plötner. *Das Serumeisen und die Eisenman-gelkrankheit.* Jena: Gustav Fischer, 1937.
22. Laurell, Carl B. *Acta Physiol. Skan* 14 supp. 46, Lund, 1947.
23. McCance, R. A. and E. M. Widdowson. *Lancet* 2: 680, 1937.
24. Michaelis, L., C. D. Coryell, and S. Granick. *J. Biol. Chem.* 148: 463, 1943.
25. Moore, C. V., W. R. Arrowsmith, J. Welch, and V. Minnick. *J. Clin. Invest.* 18: 553, 1939.
26. Moore, C. V., C. A. Doan, and W. R. Arrowsmith. *J. Clin. Invest.* 16: 627, 1937.
27. Moore, C. V., R. Dubach, V. Minnich, and H. K. Roberts. *J. Clin. Invest.* 28: 755, 1944.
28. Pederson, K. O., and J. Waldenström. *Z. physiol. Chem.* 245: 152, 1937.
29. Peterson, J. C., and S. C. Dunn. *Am. J. Dis. Child.* 71: 252, 1946.
30. Robscheit-Robbins, F. S., and G. H. Whipple. *J. Exp. Med.* 63: 767, 1936.
31. Schade, A. L., and L. Caroline. *Science* 104: 340, 1946.
32. Waldenström, J. *Om Järn och Järnterapi.* Upsala, 1944.
33. Welch, C. S., E. G. Wakefield and M. Adams. *Arch. Int. Med.* 58: 1095, 1936.
34. Whipple, G. H. and F. S. Robscheit-Robbins. *Am. J. Med. Sci.* 191: 11, 1936.
35. Yoshikawa, H., P. F. Hahn and W. F. Bale. *Proc. Soc. Exp. Biol. Med.* 49: 285, 1942.

A SYSTEMATIC TREATMENT OF COORDINATION COMPLEXES OF METALLOPORPHYRINS AND NITROGENOUS BASES

W. MANSFIELD CLARK,[1] T. H. DAVIES, C. C. PORTER, J. F. TAYLOR, JOSEPH SHACK, C. S. VESTLING AND J. R. WEISIGER

From the School of Medicine, The Johns Hopkins University, Baltimore, Maryland

Systems containing metalloporphyrins and coordinating substances may exhibit complex equilibrium states involving exchanges of electrons, protons and coordinating substances and the making and breaking of dimeric units.

Because of the complexities it is impracticable to include in this review more than a sketch of the sort of data assembled by the authors, or published by Conant *et al.*, Hogness *et al.*, Barron and others (1). Nor is it practicable to reproduce here the numerous diagrams which, with the exception of those illustrating new data, have been published.

Measurements of diffusion coefficients, and some comparative tests by means of the ultracentrifuge confirm earlier, scattered evidence that metalloporphyrins in aqueous solution form large micellae. Some of the following data indicate that within these micellae are dimeric units. The bondings in the larger aggregates and even in the dimers appear to be too weak to be reflected in those energy changes which are involved in some of the measurements of equilibrium states, but the bonding of the dimeric units is definitely reflected in some of the measurements. This poses an interesting problem in methodology.

The resulting complexity has made it advisable to preserve, so far as practicable, objective mathematical methods of analyzing data. The methods of Reed and Berkson (2) have been used in most cases. The limitations set by the precision of the data have been noted in previous publications.

Porphyrins, such as protoporphyrin IX and the corresponding meso- and hematoporphyrins and coproporphyrin I, coordinate with any one of several metals to form a class of substances which we call metalloporphyrins. The iron, cobalt and manganese complexes are subject to oxidation-reduction. We have determined the electrode potentials of a considerable variety of these systems. In some instances the more reliable potentials were obtained by use of a mediator, that is, a reversible oxidation-reduction system which affects the electrode the more definitely, keeps in equilibrium with the system under study and is present in proportions so small as not seriously to affect the state of oxidation-reduction of the principal system. Correction of the data for the interaction can be made easily when the proportions of the two systems and the characteristics of the mediator system are known. For example, previous investigators have had difficulty in obtaining steady electrode potentials with the iron protoporphyrin system. By use of indigosulfonates as mediators Shack and Clark obtained reliable data.

Whereas a ferri porphyrin contains hydroxyl ion a ferro porphyrin does not. This is evidenced by the fact that such a system at fixed degree of reduction shows a change of potential with change of pH defined by $\frac{\Delta E}{\Delta pH} = -0.06$ at 30°C. The cobalt and manganese porphyrins show no such change.

The shift of potential with pH is not attributable to a great change in the ionization constants of the carboxyl groups of proto-, meso- or coproporphyrin. This is a reasonable assumption checked by several facts. In spite of the low solubility in neutral and acid solution ferri coproporphyrin I was found by Porter to give a titration curve indicating that all the carboxyl groups are ionized in the region of pH to be considered in the following cases. Secondly, etioporphyrin, which contains no carboxyl groups, showed with pyridine the relation $\frac{\Delta E}{\Delta pH} = -0.06$.

If we assume that iron has the coordination number 6, and that four positions are occupied by the four pyrryl nitrogens of the porphyrin ring, there remain two positions to be filled. We consider it artistic and not experimentally proven that in aqueous solution water molecules can fill

[1] Paper presented by W. Mansfield Clark.

these positions. If so, the finding of evidence for coordinated $\overline{OH}$ can be explained by loss of a proton from one of the coordinated water molecules. This should be replaced at relatively low pH. Measurements of this ionization constant are difficult because of the low solubility of any one of the metalloporphyrins, but we have spectrophotometric measurements indicating that the half-transformation points are at about pH 7.4 to 7.6.

A metalloporphyrin will coordinate with any one of a large number of nitrogenous bases such as pyridine, nicotine, pilocarpine, histidine and with cyanide ions.

If compared under comparable conditions, such compounds show striking differences in the association constants. For example, the concentration of pyridine required for half-conversion of ferrimesoporphyrin to the pyridine complex may be as high as 0.3 molar whereas under comparable conditions pilocarpine will half-saturate the metalloporphyrin at 0.3 millimolar concentration. On comparing substituted pyridines Weisiger finds effects reflecting their ionization constants and also evidence of steric hindrance.

In several cases the evidence points to stepwise additions to a reduced metalloporphyrin. Although there are theoretical reasons for assuming step-wise additions to oxidized metalloporphyrin the data can be treated mathematically as if the ratio of equilibrium constants approaches zero, that is, no intermediate forms to a surely detectable extent.

In the more usual case the coordination with the reduced metalloporphyrin is the stronger. Hence the potential of the system at a given per cent reduction becomes more positive as concentration of the coordinating substance increases. This shift of potential is a reflection of the difference between the free energies of association.

Analyses of spectrophotometric data on the association between a ferri porphyrin (proto-, meso-, or copro-) and either pyridine, nicotine or pilocarpine in alkaline solution indicates that two molecules add per unit of metalloporphyrin. Apparently neither of these replaces OH, which would seem to be confirmed by the fact that the relation of electrode potential to pH is $\frac{-\Delta E}{\Delta pH} = 0.06$. In neutral solution, however, the data suggest that with the elimination of $\overline{OH}$ from a ferriporphyrin four molecules of pyridine add to the dimeric metalloporphyrin, split the dimer and form monomeric dipyridine ferri porphyrin.

In all cases it appears that the reduced ferro porphyrin is also dimeric but adds pyridine, perhaps stepwise to form monomeric complexes. These do not vary with pH.

Consequently reduction of the dimeric dipyridine ferri porphyrin should involve two electrons per unit and should be accompanied by the elimination of $\overline{OH}$ and the splitting of a dimer to a monomer. The available evidence supports this.

As was shown by Hogness *et al.*, the spectrophotometric data on the coordination of cyanide ion with ferriprotoporphyrin can be accounted for on the assumption that four cyanide ions add to this dimeric metalloporphyrin, eliminate $\overline{OH}$ and split the dimer to form monomeric dicyanide ferriprotoporphyrin. We have extended this study to show that HCN does not combine with a ferriporphyrin but only cyanide ions, and that the association is a function of both the ionization constant of hydrocyanic acid and the ionization constant of the hydroxyl of ferriprotoporphyrin.

Any one of these systems when saturated with $\overline{CN}$ gives an electrode potential invariant with pH. The curve relating potential to pH at fixed degree of reduction crosses the corresponding curve for the metalloporphyrin and indicates that when released from the necessity of competing against $\overline{OH}$ in the oxidant, $\overline{CN}$ combines more tightly with ferri protoporphyrin than with ferro protoporphyrin. The reverse is true of alkaline solutions. This suggests certain analogies with the effect of cyanide on enzyme systems. The analogy may not be good as will appear in the following remarks.

If to an aliquot of a solution of an iron porphyrin there is added $\overline{CN}$ sufficient to cause some but not an extensive formation of a complex and to another aliquot there is added pyridine or pilocarpine sufficient to cause a little but not an extensive formation of a complex, an extensive color change will take place on mixing these solutions. The extent of the formation of what appears to be a new complex is a complicated function of the concentrations of the reacting substances and of pH.

One of the more striking facts is the extremely low concentrations of the coordinating substances required to form the complex. This rules out the formation of a mixture of those complexes which are formed on the one hand with cyanide alone and, on the other, with pyridine or pilocarpine alone. We have accounted quantitatively for this phenomenon by assuming the formation

of a monomeric, monocyanide, monopyridine ferri protoporphyrin.

A first approach to an account of this fact is found on considering the competition between coordinating substances and hydroxyl ion for place in the coordination complex. Cyanide ion can compete successfully, replacing $O\bar{H}$; pyridine cannot. But on lowering pH, so that pyridine does not have to compete with $O\bar{H}$, we find pyridine behaving as does cyanide. Remove the competition of $O\bar{H}$ by means of $C\bar{N}$ and pyridine enters on about the same footing. This description is, of course, a bit crude and is resorted to only to give the over-all picture.

No account of a few of the conclusions which have been confirmed would be fair without remarking that each fresh attack reveals some hitherto unsuspected complexity. For example, some earlier work on the complexes formed with pilocarpine was repeated after a study had been made of the kinetics of hydrolysis of the lactone. The author thought he had found a striking lowering of the affinity of pilocarpine for ferri protoporphyrin due to the opening of the lactone ring. A repetition with unhydrolyzed pilocarpine under conditions such that hydrolysis could not have been of significant rapidity gave essentially the same association curve. What is remarkable is that, unlike all previous experience, the curve analyzed as if but one pilocarpine added. The only difference between these cases and those previously studied which has been detected to date is that in the latter cases the solution of ferric protoporphyrin in a borate buffer had 'seasoned' by standing for several days.

We already have published a note on the spectral changes occurring in a solution of 'heme' (ferri protoporphyrin) on standing. The spectral changes suggest alteration of degree of aggregation accompanied by slower decomposition.

REFERENCES

1. See references in the series 'Metalloporphyrins,' *J. Biol. Chem.* 135: 543, 569, 597, 623, 643, 1940; 171, 143, 1947.

2. Reed, L. J. and J. Berkson. *J. Phys. Chem.* 33: 760, 1929.

RELATION OF PHYSIOLOGICAL FUNCTION AND MOLECULAR STRUCTURE IN HEMOGLOBIN

JEFFRIES WYMAN, JR.

From the Biological Laboratories, Harvard University, Cambridge, Massachusetts

The most important physiological functions of hemoglobin in the vertebrate body are probably to provide for the transport of oxygen, to participate in the transport of carbon dioxide and to act as a buffer in maintaining constancy of the pH of the internal environment. The first of these is due primarily to the presence of four hemes in the molecule and their interaction, the third to the characteristics of the globin and the second largely to the character of the heme-globin bonding, which involves the existence of certain oxygen-linked acid groups. These groups operate as a very simple and beautiful physicochemical mechanism in accordance with the principles of acid-base equilibria. All in all, hemoglobin certainly represents one of the most striking examples, if indeed it is not *the* most striking example, known to biochemistry of the relation of structure to function at a molecular level. It is the purpose of this paper to develop this point of view a little further.

We shall begin, by way of general background, with a brief summary of existing knowledge with regard to the physical and chemical characteristics of the vertebrate hemoglobins. In all of these there are four protohemes in each molecule, and the molecular weight is close to 67,000, corresponding to an iron content of approximately 0.35 per cent. In the case of horse hemoglobin, which is probably the most completely characterized of the hemol globins, it may be deduced from physicochemica-studies that the molecule in solution may be approximated either by an oblate ellipsoid with an axial ratio of 3:1 or a prolate ellipsoid of axial ratio of 1:3, and contains about 30 per cent water of hydration (20). As between these alternatives, a clear decision is provided by the X-ray studies of Perutz and his associates (1), whose conclusions go much farther. They have inferred that in the dry crystal the molecules of horse ferrihemoglobin are right circular cylinders with convex ends, 34 A high and 57 A in diameter, consisting of two identical halves and having a two-fold axis of symmetry. They contain four equally-spaced layers of polypeptide chains. The four hemes occur in pairs on the surface of the molecule, with their planes parallel to one another and to the plane defined by the cylinder axis and the axis of symmetry. On the basis of osmotic and centrifugal studies it is known that at high dilution or in strong urea solutions, the molecule splits into two equal parts (11, 14).

As regards the ammo acid composition, the most striking feature is the almost uniquely high figure for histidine (33 residues per molecule, 16). Recent studies of Sanger and Porter[1] give information on the number of free α amino groups in several mammalian hemoglobins. In the case of the horse there appear to be six of them, all belonging to valine; in the case of cow, sheep and goat there are four, two belonging to valine and two to methionine. This illustrates the fact that there are significant differences of composition as well as of physicochemical properties between hemoglobins of different species.

With this preface, let us now consider the buffering action of hemoglobin. The total number of titratable groups in horse hemoglobin is estimated as about 178, but most of these are only active at rather extreme reactions. The groups which may be expected to be active in the physiological range are the 33 imino groups of the histidine residues and the six α amino groups of valine revealed by Porter and Sanger. Owing to the proximity of an amide linkage the pK of these latter groups might be lowered towards 7. In histidine itself the pK of the imino group at 25° is 6, and in the histidine peptides which have been studied it is somewhat less (5.6–5.8) (2, p. 85). Between pH 6 and 8 the buffering power of horse hemoglobin remains fairly constant at the high value of approximately 2.9 equivalents per heme per pH unit, which is the value characteristic of the middle of this range and is essentially the same for hemoglobin as for oxyhemoglobin (6). This accounts for the remarkably high pH stability of the blood under physiological conditions. That the acid groups involved are in fact primarily imino groups of histidine accords well with the effect of temperature on the titra-

[1] Personal communication from Professor A. C. Chibnall.

tion curve, from which an apparent heat of dissociation of approximately 6200 calories, involving about 31 groups active between pH 5.5 and 8.5, may be deduced (17). This is a value characteristic of the imidazole group of histidine These considerations offer the first instance of the important physiological rôle played by this amino acid, namely its action as a buffer.

We turn now to the question of oxygen transport. The bulk of the evidence, including that from x-ray studies, heats of oxygenation and Bohr effect, is in favor of the view that the four hemes are all alike in their reactivity with oxygen, and that each is linked with an essentially identical configuration in the globin. They are not, however, independent of one another, but are subject to stabilizing interactions such that the introduction of one oxygen molecule into the protein facilitates the introduction of a second as by a kind of decoying mechanism. This follows unequivocably from the value of $n = 2.5 - 3.0$ in the approximate Hill equation. A particularly simple and ingenious model of how that might occur has been given by Pauling and is very instructive (8). It is now clear, however, that it cannot be regarded as meeting the most exact experimental facts. Instead, it seems highly probable, on the basis of a more recent analysis (21) involving experiments on the oxygen equilibrium of hemoglobin split in strong urea solutions as well as experiments on oxidation-reduction voltages of both native and split hemoglobin, that the four hemes occur in two identical pairs with strong interactions between members of the same pair and weaker interactions between members of different pairs. The resulting effects on the steepness and sigmoid character of the oxygen equilibrium curve, which are of obvious physiological significance, are a particularly good example of the relation of structure and function in this remarkable molecule. As regards the origin of the stabilizing energy between the hemes, it may be inferred from the effect of splitting the molecule, which causes an increase in oxygen affinity, that in the case of hemes belonging to different pairs it involves primarily the unoxygenated forms. Without dwelling further on these matters, however, let us turn now to those aspects of the oxygen equilibrium which relate directly to the rôle of the oxygen-linked acid groups, groups which, as we shall see, are probably to be identified with histidine.

A knowledge of the oxygen Bohr effect, by which is to be understood the effect of pH on the oxygen equilibrium of hemoglobin, goes back nearly half a century. Actually, what Bohr and his colleagues discovered in 1904 was not the effect of pH but the more restricted effect of carbon dioxide on oxygenation of the blood. Ten years later, in 1914, Christiansen, Douglas and Haldane discovered the converse effect, the effect of oxygenation on the CO_2 equilibrium of the blood. The two effects are, of course, aspects of a single effect, and either one follows thermodynamically from the other in accordance with the general relation

$$\left(\frac{\partial \ln a_1}{\partial m_2}\right)_{m_1} = \left(\frac{\partial \ln a_2}{\partial m_1}\right)_{m_2}$$

where the a's denote activities, the m's total masses and the subscripts refer to two different components in the system. Later on it came to be realized that the two effects, referred to as the Bohr and the Haldane effects, were not specific to CO_2 but were due to the acid properties of carbonic acid, and it was L. J. Henderson who, in 1920, gave the essentially correct interpretation of it in terms of the presence of a heme-linked or, as we may say speaking in terms of function, oxygen-linked acid group in the molecule (7). It is now known that there are in reality two such groups associated with each heme, one of which is rendered stronger, the other weaker, as a result of oxygenation of the heme and that these are the same for each of the four hemes (6). At 25°C. and an ionic strength of 0.16, the pK values of these groups, which we shall refer to as 1 and 2, in horse hemoglobin are estimated as follows (19):

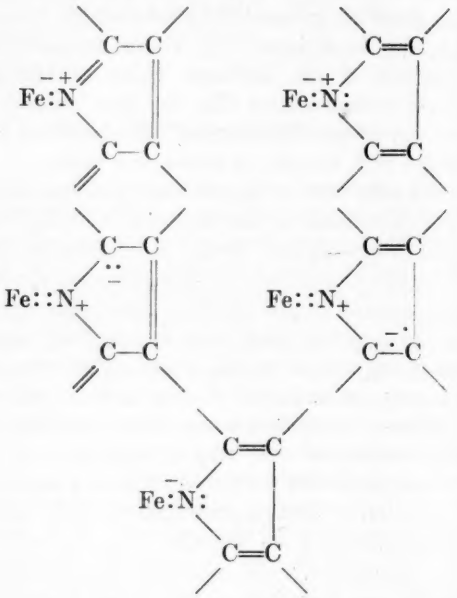
	Hemoglobin	*Oxyhemoglobin*
Group 1	7.93	6.68
Group 2	5.25	5.75

At 37°C. all figures are approximately 0.19 unit less. It is *group 1* which is accountable for the oxygen Bohr effect in the physiological range, i.e., the one originally discovered. Similar effects, involving other functions, such as oxidation-reduction, and other acid properties have since been revealed.

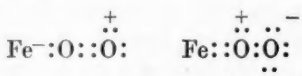
It is clearly of great interest to identify these two groups. Owing to the range in which they are active and the considerations presented earlier in regard to the buffering power of hemoglobin it is

natural to identify them as imino groups of histidine, and on the basis of structural considerations there is reason to do so, following the original suggestion of Conant (3), though it is, of course, possible that one or the other of them might be one of the free α amino groups of valine. This suggestion, at least as regards *group 1*, and probably also as regards *group 2*, is strongly supported by thermodynamic reasoning involving the heat of oxygenation of hemoglobin in relation to pH (18). Since, in the region of the Bohr effects, oxygenation is associated with the binding or dissociation of protons, the heat of oxygenation must vary with pH in accordance with the characteristic heat of dissociation of the oxygen-linked acid groups. The value of this heat may, therefore, be obtained from data on the effect of temperature on the differential titration of hemoglobin and oxyhemoglobin, though unfortunately it involves second-order effects in the observations. When this is done the answer comes out as 6500 calories for both groups, though the error in the case of *group 2*, for which the Bohr effect is reduced and which is active in a range where the protein is less stable, is greater than in the case of *group 1*.

Before considering the profound physiological importance of these oxygen-linked groups, presumably histidine, it is worth while to pause to discuss certain structural considerations. These are due mainly to Coryell and Pauling and had their origin in magneto-chemical studies on hemoglobin, free heme and certain of their derivatives. In 1936 Pauling and Coryell announced the discovery that hemoglobin was paramagnetic, but that when converted to oxyhemoglobin it became diamagnetic and that the change involved not only the protein but the oxygen molecule as well, which itself became diamagnetic as a result of the reaction (9, 10). In the case of hemoglobin it can be shown that the observed susceptibility corresponds to the presence of four unpaired electrons in each ferroheme, due to the unaltered electronic configuration of the ferrous ion (15). This can only mean that in hemoglobin the bonding of the iron to the four nitrogen atoms of the porphyrin ring, as well as to the globin, is completely ionic. On combination with oxygen there is a complete change of structure and all these bonds become covalent, like the bond formed with oxygen. As regards the covalent bonds with the four nitrogen atoms of the porphyrin ring, these may be interpreted as involving a set of resonating structures such as

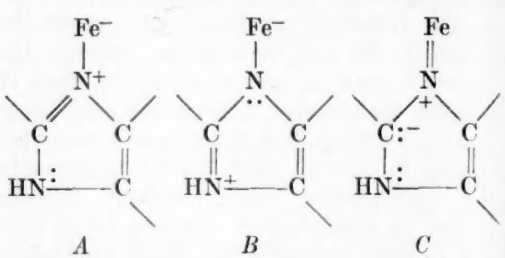


The bond with oxygen may be regarded as a hybrid between



It should be noted that in all these formulae the presence of the bivalent bonds with iron, involving electron pairs drawn from additional third orbitals, reduces the formal negative charge imparted to the iron.

In discussing the iron globin bond, Coryell and Pauling (4) assume that there is true bond formation only with the oxygen-linked acid, *group 1*. This they identify as an imidazole group of histidine. In oxyhemoglobin the bond is supposed to involve the following three resonating structures:



A *B* *C*

of which *C* plays probably a minor rôle. The acid properties of the group are due wholly to *form B*,

which may be expected to occur with about the same frequency as *A*. On this basis, by analogy with the pyridinium and free imidazolium ions, Coryell and Pauling estimate a pK for *group 1* in oxyhemoglobin of about 7. In hemoglobin, with ionic bonding, *form B* will be considerably repressed, and *group 1* should therefore be weakened, as it is known to be. It should be realized that in this heme-bonded imidazole from which a proton has been displaced in bond formation, it is the second hydrogen to which the acid properties are due.

As regards the oxygen-linked *group 2*, it is clear that this cannot be bonded with the iron in oxyhemoglobin, owing to the presence of oxygen. It follows, therefore, from the differential titration data, that it cannot be bonded with it in hemoglobin either, for bonding leads to displacement of a proton. Coryell and Pauling likewise identify this group as an imidazol group and suppose that it is simply electrostatically coordinated with the heme iron (4). In hemoglobin it is thereby strengthened beyond its condition in free histidine (observed pK 5.25 as compared with pK 6 in free histidine). In oxyhemoglobin, due to the shielding influence of the oxygen molecule, it reverts towards its condition in histidine (observed pK 5.75).

The significance of the oxygen-linked *group 1* in the exchange of carbon dioxide is most obvious when we consider the case of an individual under conditions where the respiratory quotient is 0.7. Under these conditions there is characteristically no significant change of pH of cells when venous blood is converted to arterial blood. At physiological reaction (say pH 7:3) we need only consider *group 1*. At this pH and 37°C. the difference in dissociation of this group as between hemoglobin and oxyhemoglobin amounts to almost exactly 0.6 equivalents (per heme). For every mole of oxygen taken up by the blood there must therefore be a loss of 0.6 mole CO_2 in the form of bicarbonate. In addition to this, to keep the pH of the cells constant, there must also be a loss of approximately 0.07 mole CO_2 in the form of carbonic acid, (taking the dissociation constant of carbonic acid in blood at 37°C. as 5.9×10^{-7}). The sum of these would account for very nearly the complete exchange of carbon dioxide accomplished. On the other hand, when the respiratory quotient approaches unity, both cells and plasma are found to be more alkaline in arterial than in venous blood (cells about 0.02 unit; plasma about 0.04 unit). This leads to an additional displacement of bicarbonate ion and carbonic acid, chiefly as a result of buffering due to histidine residues in the hemoglobin, which may be calculated as 0.20 to 0.25 mole. Residual amounts of CO_2 exchange may be ascribed to carbamate. The whole process represents a complicated interplay of functions in which, if our identifications are correct, a primary rôle is played by histidine and for which a detailed mechanism would be provided by the structural picture of Coryell and Pauling.

In conclusion we turn to the question of carbamate formation in the blood, which is closely related to the problem of CO_2 transport and bears upon the structural interpretations which we have just been considering. Our knowledge of this matter comes mainly from the experiments of Ferguson (5) and of Stadie and O'Brien (13). Ferguson worked with unbuffered solutions of human hemoglobin resembling whole blood; Stadie and O'Brien with similar solutions of horse hemoglobin. In the face of severe experimental difficulties both investigators showed that as pCO_2 was added there was first an increase and then a slow decline in carbamate formed, and that when the hemoglobin solutions were oxygenated, carbamate formation dropped, in some of Ferguson's experiments to one third or less of its previous value. Under the conditions of the experiments, the highest values of total carbamate were always much less than 1 equivalent per heme. (In only one observation was it as great as one third). On the basis of his results, Ferguson estimated that as much as 30 per cent of the carbon dioxide transport might be due to carbamate.

Roughton has given an interesting qualitative interpretation of Ferguson's experiments (12). In order to understand this it is necessary to recall certain general facts relating to carbamates. Carbamate formation takes place between the uncharged amino group and dissolved CO_2 (not H_2CO_3, HCO_3^-, or $CO_3^=$), and results in loss of a proton, carbamic acids being fairly strong acids (pK 6 or less):

$$R\,NH_2 + CO_2 = R\,NH\,COO^- + H^+.$$

There is no formation of zwitter ions, the NH group being an extremely weak base due to resonance phenomena. The reaction is limited to aliphatic amines and does not occur with the

imidazole group of histidine. Having regard to these considerations, Roughton arrived at the following picture. The carbamate observed by Ferguson resulted from the combination of CO_2 with free amino groups whose pK was in the neighborhood of 7. (Presumably these would have to be free α amino groups to have such a low pK.) The reason why increasing pCO_2 did not increase the amount of carbamate was because of the effect of increased acidity on the dissociation of these groups. The effect of oxygenation in decreasing carbamate formation was due to the fact that the amino groups were oxygen linked—in fact he identified them with what we have designated as *group 1* and supposed to be histidine. That oxygenation affects the reaction of this group with CO_2 was explained by steric hindrance.

This view is clearly of great interest in connection with the present discussion. In the first place it would provide a significant additional mechanism for the transport of carbon dioxide, depending merely on the introduction of oxygen into the hemoglobin molecule quite apart from changes in pH and pCO_2. In the second place, and this apparently has not been noticed previously, it would imply thermodynamically the existence of a significant new mechanism for the exchange of oxygen, for it follows necessarily that if oxygen affects carbamate formation there must be a reciprocal effect of carbamate on oxygenation. Thus, even at constant pH and pO_2, the formation of carbamate would lead to a loss of oxygen by hemoglobin. It is, however, inconsistent with our identification of *group 1* as an imidazole group of histidine and would necessitate the abandonment of the structural picture of Coryell and Pauling which seems to fit the facts so well. Nor is it so easy to devise an electronic interpretation of the Bohr effect involving *group 1* on the assumption that this is an amino group.

It is of interest to calculate the magnitude of the effect of oxygen on the carbamate equilibrium according to Roughton's hypothesis. For this equilibrium, in accordance with the considerations presented above, we may write:

$$\bar{Z} = \frac{K_1 K_2 p}{K_1 K_2 p + K_1 H + H^2}$$

in which $\bar{Z}$ is the fractional saturation of the amino group with CO_2, K_1 its acid dissociation constant and K_2 the equilibrium constant for its reaction to form carbamate. Since Roughton postulates that only one amino group per heme is involved, $\bar{Z}$ may be obtained at once by dividing the measured carbamate by the oxygen capacity of the solution. From the data given by Ferguson, including his estimated pH values (unfortunately he made no direct pH measurements) we calculate in this way K_2 to be between 7 and 8 times greater for hemoglobin than for oxyhemoglobin. This is a minimum figure; if we had assumed other groups, not oxygen linked, to have been involved as well, it would have been larger. The result is significant for it requires as a reciprocal effect that the oxygen affinity of hemoglobin must be between 7 and 8 times that of hemoglobin-carbamate. It suggests a very simple test of the hypothesis which has not yet been tried but which would avoid the severe experimental difficulties involved in direct measurements of carbamate. This would be to study the effect of pCO_2 on the amount of oxygen combined with hemoglobin at constant pH and pO_2. Actually, if the measurements were made at pH $\geqslant$ 9.5, where the Bohr effect disappears, it would be unnecessary to maintain strictly constant pH, and saturation with CO_2 should lead to a sevenfold decrease in oxygen affinity.[2] At lower pH values it would be necessary to use strongly buffered solutions. By ascribing to the amino group the pK values given above for *group 1*, we obtain by simple mass law considerations for pH 7.3 and for a constant value of pO_2 such that the solution is 80 per cent saturated with oxygen in the absence of CO_2, the ratio $\frac{dHbO_2}{dHbCO_2} \cong -2/3$. This value is relatively constant over a wide range of values of HbO_2 and $HbCO_2$. Experiments of this kind should be relatively easy and decisive.

As a final point in this discussion we offer tentatively the following explanation of the experiments of Ferguson and of Stadie and O'Brien as an alternative to that given by Roughton and one which would avoid the difficulties of his hypothesis. Whether or not it is acceptable will depend on an appraisal of the errors of the observations. Instead of thinking in terms of a single heme-linked free α amino group, let us assume that carbamate formation involves mainly ϵ amino groups, whose pK_1 values are quite alka-

[2] The fact that at such an alkaline reaction additional amino groups, say amino groups of lysine, would also take up CO_2 would make no difference.

line, say considerably greater than 8. These will not of course be oxygen linked. Then for pH $\ll$ pK$_1$, *equation 1* goes over into $\frac{\bar{Z}}{1-\bar{Z}} = \frac{K_1K_2p}{H^2}$ or, since in accordance with our hypothesis that many groups are involved $Z \ll 1$, $\bar{Z} = \frac{K_1K_2p}{H^2}$. Accordingly $\bar{Z}$ is very sensitive to pH and the effect of oxygenation in diminishing carbamate formation is to be ascribed simply to the increase of acidity accompanying oxygenation in the unbuffered hemoglobin solutions. Under the conditions of the experiments this may be estimated to be of the order of 0.1. As a test of this hypothesis we have calculated values of K_1K_2 from the data of Stadie and O'Brien, using their measured pH values, with the following results:

Hemoglobin							
p CO_2(mm)	0.7	2.3	10.2	47.4	139	271	
$K_1K_2 \times 10^{17}$	(0.25)	(0.49)	0.68	1.33	1.01	0.94	av. 0.99
Oxyhemoglobin							
pCO_2	3.6	10.6	25	53	96	177	312
$K_1K_2 \times 10^{17}$	0.66	0.57	0.59	0.77	0.97	0.80	0.50 av. 0.69

In taking the averages, the two bracketed values for hemoglobin were omitted because of high pH values, namely 8.52 and 8.15. A similar direct calculation is not possible in the case of Ferguson's data owing to lack of measured pH values. By making use of calculated pH values,[3] however, we arrive at the following results for his most complete experiment.

Hemoglobin					
p CO_2 (mm)	4.45	19.60	43.0	80.50	av. 1.69
$K_1K_2 \times 10^{17}$	1.49	1.72	1.85	1.71	av. 1.69
Oxyhemoglobin					
p CO_2 (mm)	8.28	21.30	45.30	74.20	
$K_1K_2 \times 10^{17}$	0.86	1.26	1.16	1.68	av. 1.24

It is also possible, without assuming an absolute value of K_1K_2, to construct theoretical curves giving relative values of carbamate in relation to p CO_2 on the basis of the same calculations. These curves may then be superposed to best advantage on the experimental points. This has been done for the two experiments just considered and the results are shown in figure 1. This figure also includes values of carbamate calculated from the measured pH values for Stadie and O'Brien's experiment. It will be seen that there is fair agreement between observed and calculated results, though the calculations certainly lead to somewhat smaller differences between oxygenated and deoxygenated solutions than correspond to the observations. Whether or not the discrepancy is significant is not certain, and further experiments are certainly in order.[4]

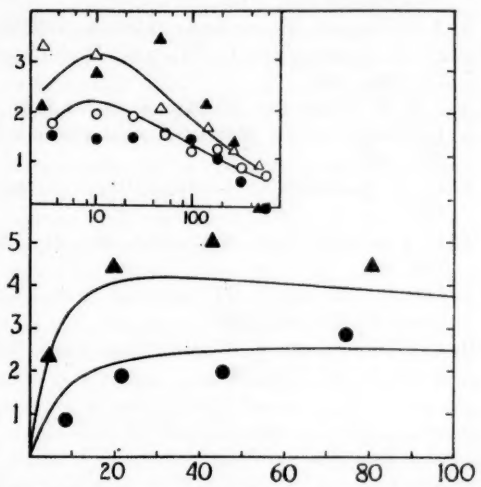


Fig. 1. CALCULATED AND OBSERVED VALUES of carbamino hemoglobin. Ordinates give $\bar{Z}$ as carbamino hemoglobin in mM per liter. Abscissae give pCO_2 (mm.H)g. Solid points are observed quantities. Curves are calculated as described in text. Open points are calculated from directly measured pH values. Triangles are for deoxygenated solutions; circles for oxygenated solutions. Main figure is for Ferguson's most complete experiment (23 x 35). Insert is for results of Stadie and O'Brien with logarithmic abscissae.

[3] In making the calculations we assume the pH of one point, take the buffer value of reduced hemoglobin as 2.9 equivalents per pH unit per heme, employ pK values of *group 1* given above, neglect the very slight effect of *group 2* in the pH range in question, take the solubility of CO_2 in the hemoglobin solutions as 2.9×10^{-5} moles liter^{-1}mm.$^{-1}$ (following Stadie and O'Brien) and (on the basis of the ionic strength of the solution) assume the first dissociation constant of carbonic acid to be 5.9×10^{-7}.

[4] Quite recently it has been pointed out by Professor Roughton (personal communication) that Ferguson's other experiments are less favorable to our interpretation and show a greater difference between K_1K_2 for hemoglobin and oxyhemoglobin than those just discussed. We have confirmed this using our calculated pH values. Thus, averaging all Ferguson's results, we obtain for hemoglobin $K_1K_2 = 1.97 \times 10^{-7}$, and for oxyhemoglobin $K_1K_2 = 1.11 \times 10^{-17}$, but the scatter is very large, extreme individual values varying by a factor of three in each series.

Whichever interpretation proves correct, however, it is clear that carbamate formation, with its possible implications regarding oxygen combination, represents another aspect of the picture we have sought to develop of hemoglobin as a highly perfected molecular mechanism designed to meet the physiological needs of the vertebrate body.

REFERENCES

1. J. Boyes-Watson, E. Davidson and M. F. Perutz. *Proc. Roy. Soc. A*, 191: 83, 1947.
2. E. J. Cohn and J. T. Edsall. *Proteins, Amino Acids and Peptides*. New York: Reinhold Publishing Corp., 1943.
3. J. B. Conant. *Harvey Lectures* 28: 159, 1932–33.
4. C. D. Coryell and L. Pauling. *Biol. Chem.* 132: 769, 1940.
5. J. K. W. Ferguson. *J. Physiol.* 88: 40, 1936.
6. B. German and L. Wyman. *J. Biol. Chem.* 117: 533, 1937.
7. L. J. Henderson. *J. Biol. Chem.* 41: 401, 1920.
8. L. Pauling. *Proc. Nat. Acad. Sci.* 21: 186, 1935.
9. L. Pauling and C. D. Coryell. *Proc. Nat. Acad. Sci.* 22: 159, 1936.
10. L. Pauling and C. C. Coryell. *Proc. Nat. Acad. Sci.* 22: 210, 1936.
11. J. Roche, G. S. Adair and M. E. Adair. *Biochem. J.* 26: 1811, 1932.
12. F. S. W. Roughton. *Harvey Lectures* 39: 96, 1943–44.
13. W. C. Stadie and H. O'Brien. *J. Biol. Chem.* 117: 439, 1937.
14. J. Steinhart. *J. Biol. Chem.* 123: 543, 1938.
15. D. S. Taylor and C. D. Coryell. *J. Am. Chem. Soc.* 60: 1177, 1938.
16. H. B. Vickery. *Ann. N. Y. Acad. Sci.* 51: 87, 1941.
17. J. Wyman. *J. Biol. Chem.* 127: 1, 1939.
18. J. Wyman. *J. Biol. Chem.* 127: 581, 1939.
19. J. Wyman and E. N. Ingalls. *J. Biol. Chem.* 139: 877, 1941.
20. J. Wyman and E. N. Ingalls. *J. Biol. Chem.* 147: 297, 1943.
21. J. Wyman. *Advances in Protein Chemistry*, 4, New York: Academic Press, Inc.

MOLECULAR OXYGEN AS A LIGAND IN METAL PORPHYRINS AND OTHER METAL-COMPLEX COMPOUNDS

L. MICHAELIS

From the Laboratories of The Rockefeller Institute for Medical Research, New York City

The iron porphyrin compounds occurring in the organism have various functions, all related to oxidation. Warburg's respiratory enzyme activates oxygen; the cytochromes transfer electrons; peroxidases activate hydrogen peroxide, catalase detroys it, hemoglobin carries molecular oxygen. Of all these functions, the most marvellous one, the one least imitable in model experiments, seems to be the ability of hemoglobin to carry oxygen, reversibly absorbing or releasing it, in response to minor variations in oxygen pressure. There is no satisfactory answer as to how this function of hemoglobin can be accounted for by its structure. The task of this paper is to discuss a number of facts, theories and hypotheses, some old and almost forgotten, others more recent, which may throw some light on this problem and may stimulate further attempts in solving this riddle. Since the whole topic is concerned with coordinative metal complex compounds, a few remarks about some general aspects of metal complex compounds will serve as an introduction.

It has been recognized for a long time that metal complex compounds may be classified in two groups according to the nature of the bond by which the ligands are bound to the central metal atom. Those two kinds were distinguished in the older German chemical literature as 'Anlagerungscomplexe' (complexes by juxtaposition) and 'Durchdringungscomplexe' (complexes by penetration). This distinction can be expressed in modern terms as follows: In one type of complex the ligands are held by electrostatic forces exerted by a positively charged metal ion upon ligands which are either negative ions with a free charge or dipoles, either permanent or induced. For instance, among the ligands frequently encountered, water is a distinct permanent dipole, ammonia is a weaker permanent dipole but it is easier to polarize. Such bonds, in Pauling's (1) nomenclature, are called ionic bonds. They are also called electrovalent, or heteropolar bonds. An example of an ionic complex is the ferric hexafluoride ion, $Fe^{III}F_6^{\equiv}$, where six fluorine ions are coordinated around one triply positively-charged ferric ion. Electrostatic forces impose no definite restriction on the number or spacial configuration of the ligands, but spacial restrictions limit this number to six which will form the corners of an octahedron. The electronic configuration of the atoms within this complex is not essentially disturbed: ferric ion, both in the state of the free (or hydrated) ferric ion, and of this complex ion, has five unpaired electrons in its incompleted shell (the 3d sub-shell). This conclusion is derived from the fact that the magnetic susceptibility is the one theoretically expected for the presence of five unpaired electrons in either case. The other type of complex compound is characterized by an essential change of the electronic configuration. In the ferrocyanide ion $Fe^{II}(CN)_6^{\equiv}$ there are no unpaired electrons at all; this complex ion is diamagnetic, although ferrous ion in the free state is strongly paramagnetic due to four unpaired electrons in the 3d sub-shell. Such compounds may be formed with such ligands which possess electrons not used for chemical bonding. The electrons of the ligands are used for the completion of the uncompleted electron shell of the metal ion so as to approach the structure of a noble gas (krypton in this case). The six ligands occupy the corners of a regular octahedron not only for spacial reasons but because the covalent bonds have the greatest stability in such a configuration, as Pauling has shown by quantum mechanical calculations. In his nomenclature all such bonds are called *covalent* bonds, or they may be called also *homopolar* bonds. In the ferricyanide ion $Fe^{III}(CN)_6^{\equiv}$ there is an odd number of electrons. Here the approach to a noble gas structure is less complete because one electron must be left unpaired whereby the paramagnetism of the ferric ion is considerably weakened indeed but not entirely abolished.

In complexes of iron it is easy to infer from magnetic measurements whether a complex belongs to the one or the other type. This is not necessarily so for all metals. So, any cupric compound will always show a magnetic susceptibility corresponding to one unpaired electron no matter what type of bond is established.

Some ligands are more inclined to form ionic bonds, others, covalent bonds. The strongly 'electronegative' fluorine ion always forms ionic bonds; cyanide ion always forms covalent bonds.

Water is usually coordinated by ion-dipole bonds, but in some cobaltic complexes by a covalent bond. Ammonia is usually bound by electrostatic bonds in cobaltous complexes, but by covalent bonds in cobaltic complexes. In hemoglobin the bonds of the iron with the N-atoms of the porphyrin ring are ionic, leaving undisturbed the paramagnetism of the central ferrous ion with its four unpaired electrons, whereas in oxyhemoglobin all bonds are covalent, destroying all paramagnetism, as Pauling and Coryell (2) have shown.

In general, the two types of bond are but extreme cases. There are many transitions which may be imagined to arise from resonance, each of the extreme configurations contributing a certain share to the actual, intermediate state. However, as regards the complex compounds of iron, cobalt and some other metals, we are faced with a peculiar situation which will now be discussed. Let us consider for instance the case of hemoglobin. How can it be explained that hemoglobin has a magnetic susceptibility as high as is to be accounted for by purely ionic bonds, with four unpaired electrons in the iron atom in its ferrous state, whereas oxyhemoglobin, which is also in the ferrous state, has no unpaired electrons at all, is diamagnetic, and shows only covalent bonds with no intermediate or resonating condition occurring? The explanation is this. In a complex iron compound the ionic form of the complex has a different number of unpaired electrons than has the covalent form. According to the rules of quantum mechanics, resonance between states possessing a different number of unpaired electrons cannot take place. On resorting to this rule, one may assume that a resonance between the two states does not occur in an iron complex compound. The molecule has to stay in one of the two possible states, whichever is the more stable one, even if the difference in stability be small. It is quite possible that even in such cases intermediate bonds do exist, as Pauling has pointed out. So, the ionic complex may be in resonance with a covalent one, not of the ordinary type but one formed by 'promoted' energy levels of some electrons in such a way that the number of unpaired electrons is not affected. Such covalent bonds, if they exist, will be weaker than regular ones; furthermore, they would permit of resonance. For this reason Pauling prefers to speak of an 'essentially' ionic, or covalent, bond in such cases. It will cause no misunderstanding if this fine distinction is disregarded in what follows. The essence of this consideration is that if there is a choice between two magnetic states, the state will be either the one, or the other, according to conditions, but never an intermediate state. If we accept this argument it will be easier to understand why such a slight perturbation as the attachment of an oxygen molecule in hemoglobin is able to throw the magnetic character of the complex entirely from one extreme to the other.

It is an essential gap in the present theory of coordination compounds that it is, in general, not possible to predict whether a ligand will form an ionic or a covalent bond. Of course, if a ligand is neither an ion nor a distinct dipole, it can form only a covalent bond. This is true, e.g., for molecular oxygen. If we accept this definition we see that it is not always true that a covalent bond is necessarily a strong or 'robust' bond.

The bond with molecular oxygen can be established in two different ways for both of which examples may be cited. The molecule O_2 contains an even number of electrons. However, the ground state of O_2 contains two unpaired electrons, which makes O_2 paramagnetic. Oxygen can be attached as a ligand either in such a way that after pairing of those two electrons, a pair of electrons is used in filling up the incomplete electron shell of the central metal atom, thus forming a 'dative bond', (as in oxyhemoglobin and in one type of the cobalt compounds to be discussed presently); or, in a binuclear complex with an O_2 bridge, each one of the unpaired electrons of O_2 may form a pair with one of the unpaired electrons of the central metal atom (as in another type of cobalt complex). In any case, oxygen, although it is itself, in the free state, paramagnetic, will, as a ligand, always diminish and usually even annihilate the paramagnetism of the complex compound.

Among other gases, CO and NO can function as ligands. In some cases, each of the three gases can be bound, as in hemoglobin. In other cases, O_2, but not CO, can be bound. CO may sometimes occupy many or all coordination places as in the metal carbonyls. Oxygen seems to be capable of coordination only in such cases where other ligands have been already bound. It is not unusual in coordinative compounds that the affinity of a ligand is strengthened if there are other ligands of a different kind already attached.

Coordination compounds containing molecular oxygen are stable compounds only in a few rare cases. Usually they are inclined to a more or less rapid intramolecular rearrangement whereby oxygen is reduced and something else, usually the central metal atom, is oxidized. One may speak

of an intramolecular redox system. This intramolecular reaction is in many cases so rapid that the existence of the original, oxygenated complex escapes direct observation and only the result of the interaction can be noticed. In some cases the intramolecular rearrangement stops with the oxidation of the metal atom. For instance, the cobaltous-bicysteine complex (3) is oxidized by oxygen to a cobaltic complex of cysteine and not readily any further. The reason is that in a cobalt complex compound, once the cobaltic state has been established it is very difficult to reduce it back to the cobaltous state. In other cases, a chain reaction is started and the metal atom may be said to activate molecular oxygen as an oxidizing agent. For instance, in the ferrous bi-cysteine complex the oxidation goes farther, cysteine being oxidized by molecular oxygen to cysteine with very little iron as a catalyst. Primarily, just as in the case of the cobalt cysteine complex, ferrous cysteine is oxidized by oxygen to ferric cysteine; then an intramolecular oxidation-reduction process leads to ferrous iron and cysteine; the ferrous iron combines with more cysteine, and a chain reaction is started. However, the very primary step in this process is the incorporation of an oxygen molecule into the ferrous cysteine complex, although this oxygenated complex is so labile that it cannot be prepared as such. As to the structure of this hypothetical labile-oxygenated ferrous bi-cysteine complex, it may contain one molecule of oxygen per one atom of iron, or, perhaps in analogy to some compounds to be discussed presently, one oxygen molecule may form a bridge between two molecules of the ferrous complex, and a binuclear complex may be formed. The hypothesis that primarily an oxygenated ferrous complex is formed is corroborated by the fact that also carbon monoxide can be bound as a ligand, and this carbon monoxide-ferrous-cysteine complex can be prepared in pure, crystalline form.

Reversible oxygenation, without any intramolecular irreversible rearrangement occurring, is encountered among the iron compounds only in hemoglobin. Cases more or less comparable are known only in cobalt compounds. The discovery of such cobalt compounds has aroused the hope of finding out under what conditions a reversible oxygenation can take place. What is known now in this field will now be briefly discussed.

The first case of this kind has been known for more than fifty years, but has been almost forgotten, or at least never been discussed as a model for oxygenation. It is concerned with the hexammine cobaltous ion, $Co^{II}(NH_3)_6^{++}$. It reacts with molecular oxygen. However, the effect of oxygen is not, or at least not primarily, the oxidation to the corresponding cobaltic compound, which is the luteo cobalt ion, but rather is the first product of interaction readily obtainable in pure crystalline form, of intensely brown color, a compound formerly called 'oxycobaltiac'. The reaction was explained by Werner (4) as follows:

$$1) \quad 2[Co^{II}(NH_3)_6]X_2 + O_2 \rightleftharpoons [(NH_3)_5Co \cdot O_2 \cdot Co(NH_3)_5]X_4 + 2NH_3$$

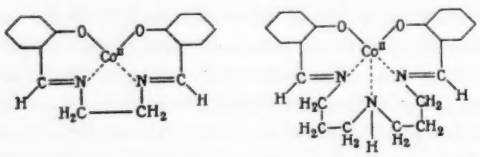
where X is an univalent anion. Since in the dissolved state the hexammine cobaltous complex, even in the presence of excessive ammonia, is always in equilibrium, at least with the pentammine, one may just as well modify Werner's formula as follows:

$$2) \quad 2[Co^{II}(NH_3)_5H_2O]X_2 + O_2 \rightleftharpoons [(NH_3)_5Co \cdot O_2 \cdot Co(NH_3)_5]X_4 + 2H_2O$$

Werner and Mylius (5) strongly suggest the reversibility of this reaction although they did not succeed in proving it experimentally for this particular case. This compound, easy to prepare and quite stable in the crystalline state, is readily decomposed by water or acids, with sudden release of all of its oxygen as gas. This fact seems in favor of assuming what we may call a cobaltous compound in the oxygenated state, although this evidence is not quite convincing because cobaltic ion, Co^{+++}, as present in cobaltic sulfate, also decomposes water with the development of oxygen, the normal potential of the couple, cobaltous ion + cobaltic ion, lying in the oxygen overvoltage range. However, as will be published separately, it is easy to arrange the experimental conditions in such a manner that the 'oxycobaltiac' can be reversibly oxygenated and de-oxygenated according to the oxygen pressure, just as with hemoglobin. In the dissolved state the effect of oxygen does not stop at this stage; rather is there very gradually established a true oxidation of the cobalt. A mixture of pentammine aquo cobaltic salts (roseo cobalt salts) and some hexammine cobaltic salts (luteo cobalt salts) is eventually established. So, the oxycobaltiac may be considered as a precursor, although a relatively stable one, to a true oxidation product. Anyhow, molecular oxygen does not directly oxidize the cobaltous state to the cobaltic by simply withdrawing one electron. Such a reaction can be

accomplished by other oxidizing agents. In this particular case a potentiometric oxidative titration of the cobaltous compound with ferricyanide does not lend itself to readily interpretable results for reasons which have been fully explained by J. Bjerrum (6). However, on using the much more chelated, strongly coordinated ethylene diamine complex instead of the ammonia complex, Bjerrum has shown that the potentiometric titration of the cobaltous complex with potassium ferricyanide yields a titration curve which precisely in every respect is characteristic of a reversible, univalent, true oxidation. Thus we can distinguish, just as in the case of hemoglobin, oxygenation as an effect of oxygen, and a true oxidation as an effect of potassium ferricyanide.

Another group of oxygenated cobaltous complexes was extensively studied by Calvin and his associates (7, 8). After Tsumaki (9) had discovered that a certain chelated cobaltous complex can reversibly absorb or release oxygen in the same manner as can hemoglobin, Calvin prepared a large number of cobaltous complexes with the



same property. They are all chelated complexes in which one or two coordination places are unoccupied. He distinguishes two types, of which the simplest representatives are shown in formulas 1 and 2. All rings established by chelation are 5- or 6-membered rings. They are prepared from *a*) a cobaltous salt, *b*) salicylic aldehyde or related compounds, *c*) ethylene diamine or (in *type 2*) a related compound. These compounds absorb molecular oxygen with the development of an intense brown color, in a reversible way. The oxygen can be pumped out or released by heating, or by acidification. The amount of oxygen absorbed depends on pressure and temperature, an equilibrium being attained as in hemoglobin. The absorption of oxygen takes place in a suitable organic solvent, or even in the solid crystalline state. The reversibility is not always quite complete. After several cycles of absorption and desorption, usually a deterioration takes place, which is at least in part some intramolecular oxidation, which however does not seem to be simply the oxidation of the cobaltous state to the cobaltic. In the 3-fluoro derivative of the complex of *type I*, no deterioration takes place even after many cycles, and this compound as a solid can be used for storage of oxygen by absorbing it from the air at room temperature and releasing it at higher temperature. Complexes of *type I* absorb one O_2 molecule for two cobalt atoms, the oxygen forming a bridge in a binuclear complex as in the oxycobaltiac mentioned before. Complexes of *type II* absorb one O_2 molecule per one cobalt atom. The magnetic properties are these. Complexes of *type II* have three unpaired electrons, just as the free cobaltous ion, and in the oxygenated state, the electrons are paired as much as possible, leaving one unpaired electron. In *type I* the cobaltous complex has only one unpaired electron, and the oxygenated complex is diamagnetic, which shows that all electrons are paired, especially also that each of the two unpaired electrons as they exist in the free O_2 molecule has paired with one electron of cobalt. There are then two types with respect to magnetic properties, to which we shall presently add a third one.

Another class of cobaltous complexes capable of reversible oxygenation has been discovered by Burk, Hearon, Caroline and Schade (10), namely, the cobalt complexes with some amino acids, especially with histidine. Here the resemblance to hemoglobin is even closer as the oxygenation takes place in an aqueous solution. The non-oxygenated complex is paramagnetic with three unpaired electrons; the oxygenated one, which is binuclear with an O_2 bridge, is diamagnetic (11), and so represents a third magnetic type, in addition to the two described by Calvin. The oxycobaltiac mentioned before also belongs to this class. In the oxygenated state it is diamagnetic, as stated already in 1911 by Feytis (12), and corroborated by the speaker, whereas all cobaltous ammines are paramagnetic with three unpaired electrons.

The reversible oxygenation of cobalt complexes with amino acids was entirely overlooked during the most interesting previous studies (13). Although oxidative changes of the cobaltous complexes on exposure to oxygen had been noticed, well defined compounds were obtained only by heating a suspension of cobaltic hydroxide with amino acids such as glycin or alanin. The surprising discovery of the various kinds of stereo isomers and their enormous rotatory power was apt to push into the background the reactions of the corresponding cobaltous compound with oxygen, seemingly resulting in an undecipherable

mixture of various compounds. This problem was studied, only recently, by Burk *et al.*, (10) especially for the case of the cobaltous bi-histidine complex. When this complex of light pink color is exposed to the air, it absorbs oxygen, the color turning intensely brown. This reaction is reversible: oxygen is released by pumping it out, or at high temperature, or on the decomposition of the complex by acids. The oxygenated cobaltous bi-histidine complex contains one O_2 for two Co atoms and is obviously a binuclear complex resembling the oxycobaltiac mentioned above. This oxygenated complex is not very stable. Gradually it undergoes an irreversible rearrangement, turning pink. According to Burk and associates, it absorbs another molecule of oxygen without there being any evidence for the oxidation of the cobaltous state to the cobaltic. Although those secondary reactions have not yet been studied in detail and no pure compounds have been reported so far, I wish to mention that according to experiments by Dr. S. Fiala in my laboratory, the first step of the irreversible reaction which establishes the light-pink compound takes place before any additional oxygen is absorbed. What is interesting at this point is only the fact that there is evidence for the existence of a reversibly oxygenated complex, which slowly undergoes an electronic rearrangement whereby its reversibility is lost.

What is true for histidine can be shown to be true also with some other amino acids, although the affinity for oxygen is much smaller. What seems to be necessary is the formation of a complex with two molecules of the amino acid, fortified by strong chelation and the formation of five- or six-membered rings. In this respect histidine is, if not unique, yet especially favorable because it provides for strong chelation with six-membered rings through both its amino-nitrogen and the ring-nitrogen.

An important problem, however, is why such a remarkable difference with respect to the stability of the oxygenated complexes prevails between iron and cobalt. No satisfactory answer can be given, yet certain facts may be correlated with this difference. There is always a reversible transition from the ferric state to the ferrous. The normal potential of the ferric-ferrous couple for various complex compounds does vary indeed, but within reasonable limits (rarely > 0.7 volts, or < -0.2 volts) (14). However, the normal potential of the cobaltic-cobaltous couple varies within enormous limits indeed. The potential of the couple, hexacyano cobaltous and cobaltic, complex is in the *hydrogen* overvoltage range; the cobaltous compound reduces water to hydrogen. The potential of the couple of the free (or hydrated) cobaltic and cobaltous ion is in the *oxygen* overvoltage range; cobaltic ion develops oxygen from water. On considering the electronic configuration, it has been made plausible by J. L. Hoard, according to a quotation in Pauling's book (1, p. 93), why cobalt complexes, provided they are covalent, should be unstable in the cobaltous form. However, it is not clear why the transition from the cobaltous state to the cobaltic is so difficult to bring about with oxygen as an oxidizing agent, in spite of the fact that oxygen so readily enters into the complex. There is obviously a high activation energy for the intramolecular oxidation reduction. For the biochemist the cobalt compounds are interesting mainly because they show the existence of oxygenated compounds and encourage the hypothesis that oxygenated iron compounds also exist, although they are just short-lived precursors of intramolecular true oxidation with the exception of the case of hemoglobin.

Finally, it is worth while discussing why molecular oxygen is usually inert and needs activation at all. Only the principle of obligatory univalent oxidation (15) can explain the inertia of molecular oxygen. This principle may be stated as follows: Whenever an oxidation takes place in solution by collision of the oxidizing molecule with the molecule to be oxidized, without the formation of a complex compound of the two, only a univalent oxidation, a transfer of a single electron, has any appreciable chance to occur. All bivalent or polyvalent oxidations proceed in univalent steps, which may more or less overlap. If the free energy involved in the first univalent step is high, this energy step may be considered as the essential part of the 'activation energy' for the overall, bivalent oxidation. In oxygen, the first step of reduction must be O_2^-, designated as 'superoxide ion' (1, 8), or, in aqueous solution, O_2H, a radical, which Latimer (16, 17) has termed 'perhydroxyl'. Estimates about the oxidation potential of the couple O_2,O_2^- have been made by Latimer (16, 17) and by Gorin (18). Such a potential, or the free energy equivalent to it, depends on the concentrations and states of the molecular species concerned. What is usually called the 'normal potential' is related to a 'standard condition', which is difficult to define for a molecular species

such as O_2^-, especially if the electron transfer takes place within the molecule of the complex compound. For this reason the normal potential of the couple O_2,O_2^-, as calculated by the authors mentioned, is not of much avail. However, qualitatively at least the principle may be stated as follows. The energy necessary to bring about the first step in the reduction of oxygen, $O_2 + \epsilon \rightarrow O_2^-$ is rather high for the free oxygen molecule. When O_2 is a constituent of an oxygenated metal complex, this energy may be quite different. In an oxygenated ferrous complex, the energy of the intramolecular reaction $(O_2,Fe^{++}) \rightarrow (O_2^-,Fe^{+++})$ or, if the complex be binuclear, $(Fe^{++}, O_2,Fe^{++}) \rightarrow (Fe^{+++}, O_2^-,Fe^{+++})$ is small and the process will readily proceed. Whereas the energy of the analogous intramolecular reaction with cobalt is very high, the reaction will not readily proceed in spite of the fact that in many cobalt complex compounds the cobaltic state, once it is established, is much more stable than the cobaltous one, and therefore the oxidation of most cobaltous complexes to the cobaltic state can be brought about with great ease with other oxidizing agents which do not require much activation energy, such as potassium ferricyanide.

Among the various, mostly unknown factors which may influence the activation energy of molecular oxygen, it may be justified to indulge in the following speculation. If an oxygen molecule forms a bridge in a binuclear complex, there is some chance that if one electron happens to jump from one ferrous iron to O_2, almost simultaneously another electron may jump from the other ferrous atom also to O_2. In this case the oxidation level of O_2^-, or H_2O_2, can be reached with much greater ease, since H_2O_2 is very much more stable than O_2H. In the same manner, H_2O_2, as a bridge, could be reduced directly to H_2O. One may wonder whether the function of the specific protein, globin in hemoglobin, is to fix the position of the heme molecules and keep them apart from each other so that an oxygen molecule attached to one of them can never form a bridge to another. This would be at least one factor, among others unknown, in stabilizing oxyhemoglobin and prevent oxygen from oxidizing it to methemoglobin, in contrast to all hematochromogens, including denatured hemoglobin, which are readily oxidized to the ferric state by molecular oxygen.

REFERENCES

1. Pauling, Linus. The *Nature of the Chemical Bond*. New York: Oxford University Press, 1939.
2. Pauling, L. and C. D. Coryell. *Proc. Nat. Acad. Sc.*, U. S., 22: 159, 1936.
   Schubert, M. P. *J. Am. Chem. Soc.* 53: 3851, 1931; 54: 4077, 1933; 55: 4563, 1933.
   Theorell, Hugo. In *Adv. in Enzymology*. 7: 265, 285, 1947.
3. Michaelis, L. and M. P. Schubert, *J. Am. Chem. Soc.* 52: 4418, 1930.
   Michaelis, L. and C. V. Smythe. *J. Biol. Chem.* 94: 329, 1931.
4. Werner, Alfred. *Ann.* 375: 1, 1910.
5. Werner, Alfred and Albert, Myllius. *Z. anorg. Chem.* 16: 245, 262, 1898.
6. Bjerrum, Jannik. *Metal Ammine Formation in Aqueous Solution: Theory of the Reversible Step Reactions*. Copenhagen, 1941. Pp. 224, 245.
7. Calvin, M., R. H. Bailes and W. K. Wilmarth. *J. Am. Chem. Soc.* 68: 2254, 1946.
8. Calvin, M. and C. H. Barkelew. *J. Am. Chem. Soc.* 68: 2267, 1946.
9. Tsumaki, T. *Bull. Chem. Soc. Jap.* 13: 252, 1938.
10. Burk, D., J. Hearon, L. Caroline, and L. A. Schade. *J. Biol. Chem.* 165: 723, 1946.
11. Michaelis, L. *Arch. of Biochem.* 14: 17, 1947.
12. Feytis, E. *Compt. Rend.* 152: 708, 1911.
    Gmelin, Handbuch der anorganischen Chemie, 8th Edition. System Nu. 58, Teil B: Die Ammine des Kobalts. Berlin, Verlag Chemie, (1930)
13. Ley, H. and H. Winkler. *Ber.* 42, 3894, 1909; 45: 372, 1912.
14. Michaelis, L. and E., Friedheim. *J. Biol. Chem.* 91: 343, 1931.
15. Michaelis, L. *Ann. New York Acad. Sc.* 11: 39, 1940.
    Michaelis, L. *J. Am. Chem. Soc.* 63: 2446, 1941.
16. Latimer, W. M. The *Oxidation States of the Elements and Their Potentials in Aqueous Solutions*. New York: Prentice-Hall, 1938.
17. Latimer, W. M. *Oxidation Potentials*. New York 1938.
18. Gorin, M. H., *Ann. New York Acad. Sc.*, 11: 123, 1940.